Lari Järvenpää

Production Planning of a Wind Farm Based on Wind Speed Forecasting

Faculty of Electronics, Communications and Automation

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 12.4.2010

Thesis supervisor:

Prof. Heikki Koivo

Thesis instructor:

M.Sc.(Tech.) Arto Tuominen

Aalto University
School of Science
and Technology

AALTO UNIVERSITY
SCHOOL OF SCIENCE AND TECHNOLOGY

ABSTRACT OF THE
MASTER'S THESIS

Author: Lari Järvenpää

Title: Production Planning of a Wind Farm Based on Wind Speed Forecasting

Date: 12.4.2010 Language: English Number of pages: 13+93

Faculty of Electronics, Communications and Automation

Department of Automation and Systems Technology

Professorship: Systems Technology Code: AS-74

Supervisor: Prof. Heikki Koivo

Instructor: M.Sc.(Tech.) Arto Tuominen

In recent years, the use of wind power has expanded significantly in many countries and shall continue on that track in the future as well. In the everyday operation of wind power producers, predicting the future output of wind power is of key interest. The structure of electricity markets imposes upon the producer to forecast the future production level, which thus forces the producer to be subject to possible deviations. These deviations lead to economic losses; therefore there is a strong need to predict as accurately as possible.

After briefly introducing wind power in general as well as explaining the structure of electricity markets, this thesis presents some of the current state-of-the-art wind power prediction models and uses as a reference the current forecasting performance from a wind farm case study. The reference model is based directly on a turbine power curve. Based on these results it then develops forecasting models for making short-term (up to two days) predictions of wind power output. The main focus is on advanced artificial intelligence-based autoregressive models enhanced with numerical weather prediction information.

The prediction of wind speed and production is rather intractable by nature; therefore there cannot be any model that would lead to zero errors of prediction. For evaluating this omnipresent uncertainty, methods for estimating prediction errors which would lead to more efficient wind power trading and management are also presented and tested.

The performance of the implemented models is measured and compared to the reference model. The results show that the use of an advanced forecasting model decreases the prediction errors and thus also the related economical losses. The results have a general value; nevertheless they, as well as the model configurations presented here, have to be considered as case-specific.

Keywords: wind farm, wind power prediction, wind speed forecasting, prediction uncertainty, electricity markets, support vector machine

AALTO-YLIOPISTO
TEKNILLINEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Lari Järvenpää

Työn nimi: Tuulivoimapuiston tuotannonsuunnittelu tuuliennusteiden pohjalta

Päivämäärä: 12.4.2010 Kieli: Englanti Sivumäärä: 13+93

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Automaatio- ja systeemitekniikan laitos

Professuuri: Systeemitekniikka Koodi: AS-74

Valvoja: Prof. Heikki Koivo

Ohjaaja: DI Arto Tuominen

Viime vuosina tuulivoiman käyttö on lisääntynyt merkittävästi monessa maassa ja saman trendin oletetaan jatkuvan myös tulevaisuudessa. Tulevan tuotannon ennustaminen on tuulivoiman tuottajille ensiarvoisen tärkeää, sillä sähkömarkkinoiden rakenne pakottaa tuulivoimatuottajat ennustamaan tulevan tuotantonsa etukäteen. Tuulivoimatuottajien on siis hyväksyttävä, että mahdolliset ennustepoikkeamat johtavat taloudellisiin tappioihin. Tämän vuoksi tuottajilla on tarve tehdä ennusteensa mahdollisimman tarkasti.

Tämä diplomityö esittelee ensin yleisesti tuulivoimaa ja sähkömarkkinoiden rakennetta. Tämän jälkeen työssä käydään läpi useita viimeisimpiä tuulituotannon ennustemalleja. Lisäksi työssä esitellään erään tuulivoimapuiston nykyinen suoraan tuuliturbiinin tehokäyrään perustuva ennustemenetelmä ja arvioidaan sen tarkkuutta. Näihin tuloksiin perustuen työssä kehitetään ennustemalleja, joiden avulla voidaan tehdä lyhyen ajan (alle kahden päivän) tuulituotannon ennusteita. Pääasiallisena mallien luokkana ovat kehittyneet tekoälyyn perustuvat autoregressiiviset mallit, joita täydennetään numeeristen sääennusteiden informaatiolla.

Tuulennopeuden ja tuotannon ennustaminen on kuitenkin varsin hankalaa, joten ei ole olemassa mallia, jonka avulla päästäisiin täydellisiin ennusteisiin. Jotta ennusteen epävarmuutta voitaisiin arvioida ja tuulituotantoa myydä ja hallita tehokkaammin, tämä työ esittelee myös menetelmiä ennustevirheiden estimoimiseksi.

Käyttöönotettujen mallien suorituskykyä mitataan ja verrataan verrokkimalliin. Tulokset osoittavat, että kehittyneiden ennustemallien avulla ennustevirheet ja näistä johtuvat taloudelliset tappiot pienenevät. Tässä työssä esitellyillä tuloksilla ja mallirakenteilla on yleistä arvoa, mutta niitä pitää tulkita tapauskohtaisesti.

Avainsanat: tuulivoimapuisto, tuulivoimaennustaminen, tuulennopeuden ennustaminen, ennusteen epävarmuus, sähkömarkkinat, tukivektorikone

Preface

Yes, 'n' how many times can a man turn his head,
Pretending he just doesn't see?
The answer, my friend, is blowin' in the wind,
The answer is blowin' in the wind.

– Bob Dylan

This LaTeX-written Master's thesis is made as an assignment for the power procurement department of the Finnish power producer Pohjolan Voima. Today, the company possesses the largest share of wind power production resources in Finland; in addition it also has vast wind power investment plans sighted for the future.

To the best of my knowledge, this thesis is the first one written specifically for the needs of a Finnish wind power producer. In this way, the thesis for its part collaborates to the scientific work done in the field and serves as a guidebook for other wind power operators as well. My hope is that this continues to be a guide in the future as well, as a high level of generality is preserved throughout the thesis.

There are several persons to whom I wish to express my sincere gratitude. First, I would like to thank my instructor Arto Tuominen, whose expertise especially in the area of electricity markets has been of great value in all phases of the work. Secondly, my supervisor professor, Ph.D. Heikki Koivo deserves the highest praise for being very encouraging, and providing me with new ideas and thoughts. I would also like to thank D.Sc. Vesa Hasu for providing me with his expertise on meteorology. My superior Jussi Hintikka and my colleagues Mikko Rajala, Anu Hyvönen, Lauri Luopajärvi and Raine Laaksonen also deserve my special recognition. Working with such enthusiastic people has been a true pleasure for me.

Finally, I wish to express my humble gratitude and respect towards my dear family and friends. It is with your love, care and support that I have been able to come this far. I am convinced that you will keep me going in the future as well, wherever I will be.

Helsinki, April 12, 2010

Lari Järvenpää

Contents

Symbols and abbreviations

Symbols

w	wind speed
d	wind direction
E	energy
m	mass
ρ	air density
A	area (of a wind turbine rotor plane)
L	length
ϖ	pressure
$\dot{y}$	time derivative of y
η	efficiency
y^*	optimal value of y
ν	Weibull distribution shape parameter
ς	Weibull distribution scale parameter
a	price area
h	hour
$D(x)$	inverse demand function, gives price at given volume
d_a^h	equilibrium demand
$S(x)$	inverse supply function, gives price at given volume
s_a^h	equilibrium supply
$t, \tau \in \mathbb{T}$	time instants, set of time instants
y_t	value of variable y at time instant t
R_t	power producer revenue
Π	power production random variable
Λ	price random variable
$\pi_{d,t}$	power committed to day-ahead market
$\pi_{i,t}$	power committed to intra-day market
$\lambda_{d,t}$	(spot) price in day-ahead market
$\lambda_{i,t}$	(average) price in intra-ahead market
IC_t	imbalance cost
$\pi_{g,t}$	generated power
δ_t	realized production deviation
$\lambda_{down,t}$	down-regulation price
$\lambda_{up,t}$	up-regulation price
$\lambda_{+,t}$	down-balancing price
$\lambda_{-,t}$	up-balancing price
$r_{+,t}$	ratio of down-balancing to spot price
$r_{-,t}$	ratio of up-balancing to spot price
$f_Z(z)$	probability density function of random variable Z
$F_Z(z)$	cumulative distribution function of random variable Z
$\omega_{+,t}$	probability of system needing down-regulation

$\omega_{-,t}$	probability of system needing up-regulation
$\vec{F}$	turbulent and viscose friction
Q	non-adiabatic net source of heat
$\vec{V}$	wind velocity
$\vec{g}$	gravitation
$\vec{\Omega}$	angular velocity
T	temperature
R	gas constant
S	water steam net source
c_p	specific heat of dry air at constant pressure
κ	water vapor mixing ratio
$\hat{y}_{t\|\tau}$	predicted value of y at t, issued at τ
l_x	number of versions of variable x (in either past or future)
ξ	external explanatory variable
$\hat{\chi}_{t\|\tau}$	predicted external explanatory variable
P	total production of wind farm
p	relative total production of wind farm, in the range [0,1]
P_{nom}	nominal (rated) power of wind turbine
P_{min}	minimum production of wind turbine
q	number of wind turbines online
Φ	information set
$\aleph$	parameter set
ϑ_M	performance measure w.r.t. metric M
ι_ϑ	improvement w.r.t. performance measure ϑ
σ	standard deviation
bias	bias component of RMSE
sdbias	standard deviation bias component of RMSE
disp	dispersion component of RMSE
$r_{x,y}$	correlation of x and y
e	error
$\hat{e}$	residual
$\varphi(B)$	AR lag polynomial
$\theta(B)$	MA lag polynomial
ψ	wavelet (mother) function
μ	weight
ω	probability
$\beta(\omega)$	quantile regression coefficient
$Q(\omega, x)$	quantile function, given probability ω and regressors x
b	bias
$\wp$	regularization constant
$\mathscr{L}$	Lagrangian
α	Lagrange multiplier
$\mathscr{J}$	cost function
$K(x, y)$	kernel function

ℓ	kernel bandwidth
$\tilde{\hat{w}}$	wavelet-AR(X) forecast of wind
γ	forgetting factor

Operators

$\dfrac{\mathrm{d}}{\mathrm{d}x}$	derivative with respect to variable x
$\dfrac{\partial}{\partial x}$	partial derivative with respect to variable x
$\int_A f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x$	integral of $f(x)$ over interval $A = [a, b]$
$\sum_i$	sum over index i
∇f	gradient of scalar field f
$\nabla \cdot \vec{F}$	divergence of vector field $\vec{F}$
$\vec{A} \times \vec{B}$	cross product of $\vec{A}$ and $\vec{B}$
$Bx_t = x_{t-1}$	lag
$\Delta^d = (1 - B)^d$	difference ($d \in \mathbb{Z}_+$)
$E[X] = \int_\Omega X dP$	expectation of X (integral of X over sample space Ω w.r.t. probability measure P)
$Var[X] = E[(X - E[X])^2]$	variance of X
$\wedge$	and
$\vee$	or
$\bar{f}$	complex conjugate of f
inf	infimum, greatest lower bound
max / min	maximize / minimize
arg max / arg min	argument of the maximum / minimum
$\|\|x\|\|_2$	Euclidean norm

Abbreviations

w.r.t.	With Respect To
NWP	Numerical Weather Prediction
(F)WPP	(Finnish) Wind Power Producer
TSO	Transmission System Operator
ECMWF	European Centre for Medium-range Weather Forecasts
AROME	Application de la Recherche à l'Opérationnel à Méso Echelle
HIRLAM	High Resolution Limited Area Model
WAsP	Wind Atlas Analysis and Application Program
SCADA	Supervisory Control And Data Acquisition
MOS	Model Output Statistics
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
SVM	Support Vector Machine
LS-SVM	Least-Squares Support Vector Machine
FF	FeedForward
BP	BackPropagation
RBF	Radial Basis Function
AR	AutoRegressive
ARX	AutoRegressive with eXogenous inputs
MA	Moving Average
ARIMA	AutoRegressive Integrated Moving Average
NARX	Nonlinear AutoRegressive with eXogenous inputs
EWMA	Exponentially Weighted Moving Average
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
AIC	Akaike Information Criterion
PDF	Probability Density Function
CDF	Cumulative Distribution Function
NPRI	Normalized Prediction Risk Index
QR	Quantile Regression
MF	Membership Function

List of Figures

List of Tables

1 Introduction

Life in the form we know it today would hardly be possible without plentiful resources of energy. It is needed everywhere: in transportation, heating, fabrication of vast amount of different commodities; in work and leisure, during day and night. One of the most important forms of energy we use nowadays is electricity. Although there are signs of electricity use being procyclical with the latest economic downturn, the general tendency is clear: we constantly need more electricity in order to live a life according to acceptable standards by which the Western world has become accustomed to. Also, the amount of individuals desiring such a level and quality of life is ever-growing.

The art of producing energy up to these days has been rather straightforward. Man knows well how to design, build and operate power plants that use different types of fuels in order to turn their energy into electricity which feeds our grids. There are however two very significant problems: firstly, the demand of electricity continues to increase; this substantiates the requirement to utilize various forms of energy to produce electricity. Secondly there is an increased awareness of the harmful effects to the environment, which former energy production methods have caused. This is in and of itself a troubling aspect for all of mankind. Thus the challenge is serious: in order to preserve the ever-increasing growth of acceptable living conditions for future generations, we must commence to harness forms of energy that are environmentally sound. New, environmentally friendly forms of energy are called *renewable*.

Wind power is one of the most promising forms of renewable energy. It provides a very attractive way to produce electricity as it is cost free. It would at first seem that mankind throughout the years has been quite blind to a viable energy solution, which would solve the problems of expensive fuels and environmental damages caused by them. Sadly, the story is only half-told as there are significant challenges related to the use of wind power as a primary source of fulfilling our energy needs.

Some of the challenges are purely technological: for example it is an engineering challenge to build a reliable and performing wind turbine for conditions which may indeed be very harsh. Other challenges are purely economic, albeit the cost of wind is free: for example turbine production is costly. There is thus a need for ensuring that the investment pays itself off with a reasonable profit margin, which is linked to the selling price of electricity. In many countries this margin is ensured by wind power subsidies. Lastly, other economic challenges lie on the foundations of electricity markets, from which wind power producers obtain their revenues.

The subject of this thesis, production planning of a wind farm, must be understood in a short sense: the undoubtedly subtle arts of investment analysis of wind power feasibility or technical turbine characteristics are not discussed within this thesis. We only take these as a given and look further: in order to such vast investments on wind farms to be economically profitable, proper management of wind resources is essential. An important part of such management is estimating the wind power output in the future.

The main questions this thesis tries to analyze are the following: firstly, we seek for the rationale of making power forecasts; secondly, we look at the current prediction procedures and ask whether there is a need for building a more sophisticated forecasting model. Thirdly, we desire to answer the question of how to design a good wind power forecasting model. Lastly, the results are questioned: does there exist significant improvement worth the extra effort of building a new model as well as new (perhaps costly) operational procedures?

For reasons that become clear later on in this thesis, it turns out that making inaccurate wind power predictions has a great impact on revenues wind power producers earn. This is why the art of wind power forecasting, as well as various wind power trading strategies, have gathered a lot of attention in recent years. There truly are strong economic incentives in designing accurate and reliable forecasting models for operational use. The list of previous theoretical work in this field of science is rather long, but some classics, and some old state-of-the-art models, are presented in [29]. A rather thorough, if not exhaustive, review on different modeling approaches is presented later on. Such a list serves as the theoretical basis of this thesis.

Based on preliminary knowledge of electricity market principles and forecasting models, the theory is put into practice as the actual goal of this thesis is to design and implement a forecasting model, which is then tested on real wind and production data. This forms the empirical part of the thesis and is of real value for all wind power producers who wish to build a forecasting model. To facilitate this purpose, general guidelines and design paradigms are presented, as well as a mathematical description of the chosen forecasting models.

The methods and models presented in the literature, and here, naturally have their individual limitations. The old wisdom known as the *Occam's razor* of a good model capturing something essential of the phenomenon in question, but not being overly complex, is true here as well. The available data, time and computational power all pose their limits on the feasibility of different approaches both in wind power forecasting and trading strategies. It must furthermore be accepted that the research aiming at improving the forecasting results is an ongoing process. The aim of this thesis, as discussed, is therefore to present some state-of-the-art methods and test them on real data. The testing is however subject to the data itself due to the fact that varying conditions significantly alter the results. Therefore the 'best model' used in this thesis applies only for the specific wind farm used in the testing.

The remainder of the thesis is organized as follows:

- Section 2 briefly discusses characteristics of wind, and wind power in general.
- Section 3 provides an overview on electricity markets, concentrating on the Nordic market NordPool. Also strategic bidding methods are presented. This Section is fundamental in understanding the need for making power predictions.
- Section 4 presents literature findings on wind power forecasting methods. The

role of this Section is central in this thesis as it serves as a reference for scientific work on wind power forecasting.

- Section 5 makes a study on wind power forecasting using real wind farm data. Based on these findings, a need for designing a new forecasting model is pointed out.
- Section 6 introduces the implemented advanced wind power forecasting models with their mathematical details. These models are based on earlier literature findings.
- Section 7 presents the results of the work.
- Section 8 summarizes the work and presents some guidelines for future research.

Finally, we would like to note that the terms *wind power* and *wind energy*, as well as *farm* and *park* referring to wind turbine clusters, are used interchangeably in this text. The written computer code used in evaluating different wind power modeling approaches is not attached to the thesis; however it is available upon request.

2 Wind power

In this Section we investigate the history and perspectives of wind power and the basic physical principles guiding the use of wind as a source of energy. We also point out some of the significant problems and difficulties related to wind power generation, especially in selling it to the markets. Finally we discuss wind characteristics and the way it is transformed to power.

2.1 Wind as a source of electricity

Wind is one of nature's most copious and intermittent sources of energy. The idea of using wind as a source of energy is very old, but it did not become economically sustainable until the 1980s. According to [6], at that time in California a combination of state and federal energy investment tax credits helped to stimulate a rapidly expanding market for wind power. During the period between 1980-1995, approximately 1700 MW of wind power capacity was erected. Concurrently in the early 1990s this same boom occurred in Germany, where 200 MW of wind energy capacity was fabricated annually. This was possible because of significant development of concepts related to wind turbines carried out by manufacturers. Following Germany, Spain and Denmark experienced a huge expansion in the use of wind power.

The perspectives of wind power are very encouraging in the future. With the endorsed EU target of reducing the total greenhouse gas emissions of the EU as a whole by 20% below their 1990 level by the year 2020 [17], the countries are facing a tremendous challenge. Producing electricity is one of the main components of greenhouse gas emissions, so guiding the power production business towards cleaner production means is essential. Increasing the share of renewables in the electricity production mix is one of the ways to achieve this common goal. The EU legislation is in line with this goal as it requires a 20% renewable energy share by the same year 2020.

The endeavor towards cleaner electricity production forms has created an ever-growing market for manufacturers of wind turbines. The same applies to auxiliary producers of services related to the production, for which maintenance is a good example. The technical challenges related to wind power production are far from being resolved: examples of these are the use of *off-shore* wind turbines, where winds normally blow stronger, and also in arctic regions. The size of turbines is also growing, compared with the current standard of 3–5 MW per windmill. It is estimated in the future this may increase to 10 MW [6]. There is also a great deal of research concerning the optimal use and management of wind farms.

Within every electrical generating process the main characteristics are as follows: Regardless of its type, a fuel is used to turn a turbine, which in turn drives a generator, feeding the electricity grid. In wind power this fuel is naturally the wind; however there is a significant distinction between conventional electricity generation and wind power. The latter is not controllable by human intervention; rather it is

dictated by meteorological facts present during any given moment. Consequently, there is little a wind power producer can do to control the operation of wind turbines fed by this omnipresent, but rather intermittent fuel. It can only switch the production off completely, and naturally this is seldom the desired type of action sought. The crude fact is that no amount of money or effort can buy additional wind once a wind power turbine, not to mention a whole farm, is built, nor to schedule the production in any other way.

Wind turbines convert the kinetic energy of wind into mechanical energy by an energy conversion process. Ideally, this process can be described by a characteristic power curve of a wind turbine. Such a curve is usually supplied by the wind turbine manufacturer. Figure 1 gives an illustrative example of a power curve. One can see that below a so called *cut-in speed* the generated power is zero, as it is above a *cut-off (shut-down)* speed also. A wind turbine always has in addition a *rated speed*, above which the power output remains constant.

Whereas some friction exists within the wind turbine machinery, it cannot operate before a certain wind speed is attained. On the other hand, the turbine cannot infinitely take power from the wind; a fact which will be addressed later on. In addition, when the nominal or rated power of a turbine is about to be exceeded, turbine control ensures that the power is limited to the rated limit. When strong winds are present the turbine is stopped in order to prevent breakages. Depending on the turbine these operational limits are usually in the order of 2–4 m/s and 25–30 m/s respectively.

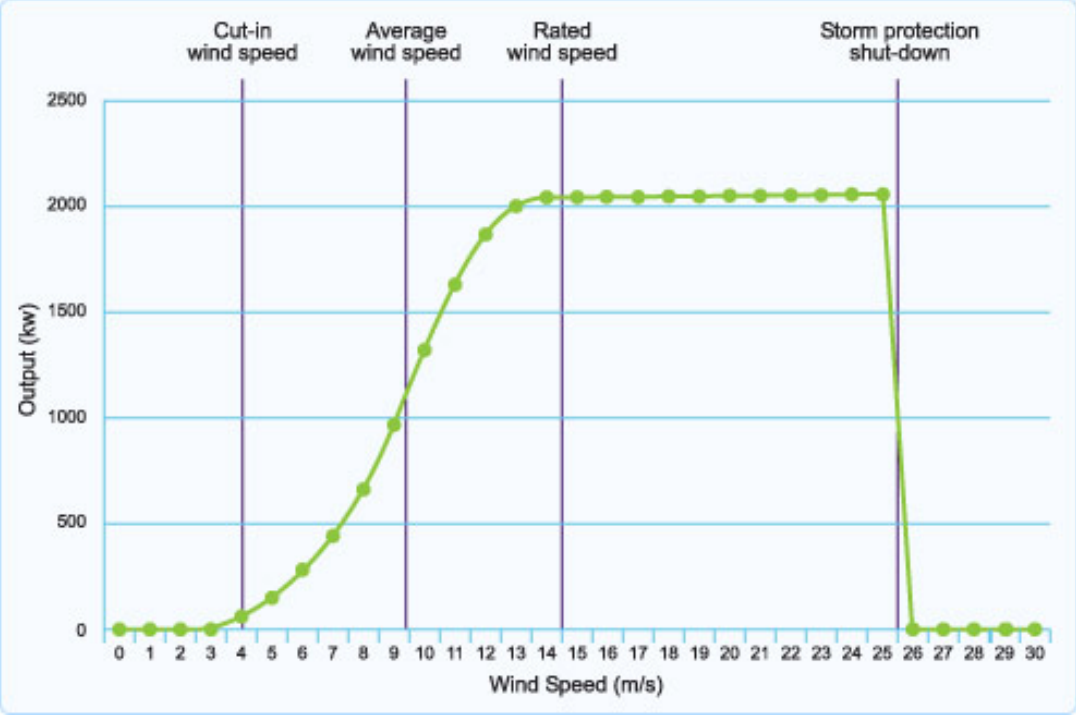

Figure 1: An example of a wind turbine power curve. Source: www.pfr.co.uk

The wind turbine power curve can be theoretically derived as follows:

The kinetic energy of a particle with a mass m moving at velocity w is

$$E = \frac{1}{2}mw^2 \tag{1}$$

Furthermore, for air density ρ and volume V we have

$$m = \rho V \tag{2}$$

and

$$\Delta V = Aw\Delta L \Leftrightarrow \dot{V} = Aw \tag{3}$$

where A is the area of the wind turbine rotor swept by the wind and V is the corresponding volume of a wind packet of length L.

By inserting (3) into (2) we have

$$\dot{m} = \rho Aw \tag{4}$$

and finally by plugging (4) into (1) the formula for the ideal wind power

$$\dot{E} = P = \frac{1}{2}\rho Aw \cdot w^2 = \frac{1}{2}\rho Aw^3 \tag{5}$$

Investigating (5) one can see that the power wind generates is essentially a cubic function of wind velocity. This explains the sharp rise between the *cut-in speed* and *rated speed* limits seen in Figure 1.

This is, however, only the ideal power of wind blowing at velocity w. Figure 2 shows the air flows before and after the turbine. By calculating those flows it can be shown that there exists a theoretical limit for the energy extracted from the wind, which is known as the *Betz coefficient*. It can be derived as follows:

Following the notation in Figure 2, the power P developed by wind turbine calculated by the pressure ϖ change is

$$P = (\varpi_1 - \varpi_2)A_t w_t \tag{6}$$

The continuity of volume flow is written as

$$A_u w_u = A_d w_d = A_t w_t \tag{7}$$

Now, as $P = Fw$, we have the force F exerted on the turbine, which equals the change of momentum between the far downstream and upstream air flows:

$$(\varpi_1 - \varpi_2)A_t = \rho A_u w_u (w_u - w_d) \tag{8}$$

We also have the Bernoulli equation for the downstream and upstream flows:

$$\begin{aligned} \varpi_\infty + \frac{1}{2}\rho w_u^2 &= \varpi_1 + \frac{1}{2}\rho w_t^2 \\ \varpi_\infty + \frac{1}{2}\rho w_d^2 &= \varpi_2 + \frac{1}{2}\rho w_t^2 \end{aligned} \tag{9}$$

Now plugging (9), (8) and (7) together we obtain

$$(\varpi_1 - \varpi_2) = \rho\frac{A_u}{A_t}w_u(w_u - w_d) = \rho w_t(w_u - w_d) = \frac{1}{2}\rho(w_u^2 - w_d^2) \tag{10}$$

This hence implies that the wind velocity through the turbine disc plane is the mean of downstream and upstream wind velocities:

$$w_t = \frac{1}{2}(w_u + w_d) \tag{11}$$

Finally, the ratio, known also as the *power constant*, between the ideal and the theoretically limited power from (6), (8) and (10) is

$$\eta = \frac{P}{\frac{1}{2}\rho w_u^3 A_t} = \frac{1}{2}\left(1 - \frac{w_d}{w_u}\right)\left(1 + \frac{w_d}{w_u}\right)^2 \tag{12}$$

Derivating (12) with respect to $\frac{w_d}{w_u}$ and setting it equal to zero, we have $(\frac{w_d}{w_u})^* = \frac{1}{3}$ and hence the upper limit for the efficiency η:

$$\eta^* = \frac{16}{27} \approx 0.593 \tag{13}$$

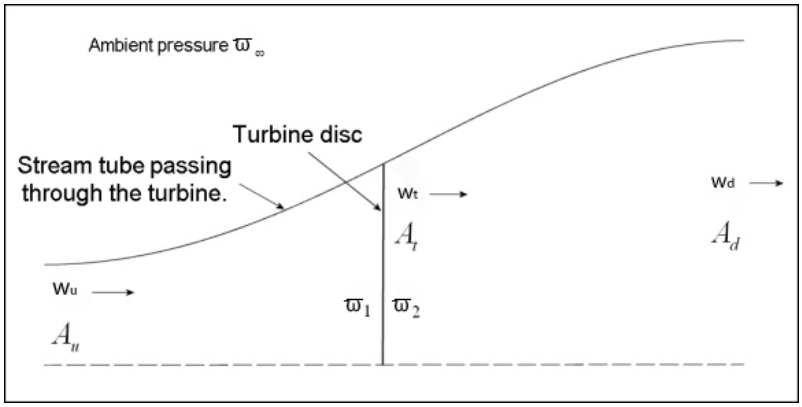

Figure 2: Flow diagram of wind before and after turbine rotor disc. Modified from: www.wind-power-program.com

This is the theoretical *Betz limit* which cannot be overcome by any kind of wind turbine technical design. In practice, the power constant is even lower due to losses

in e.g. the wind turbine generator and gearbox. A good turbine nowadays has a power constant of around 0.35 [64].

As a wind turbine is designed with the purpose of producing maximum power for a given wind speed, it has to turn its rotor towards the strongest wind all the time. Furthermore, it has to ensure that the rated maximum power is not exceeded. Controlling methods have to be applied in order to achieve these goals effectively. There are two main paradigms in wind turbine controlling: the *pitch* and *stall* control. For an overview on these methods we refer to [6].

2.2 Wind power intermittence

The question facing every WPP considering building a wind farm somewhere is: does there exist enough wind so that the production will be economically viable? As already stated, the wind turning the turbine rotor is the single most important factor in designing wind power production. How can one characterize the wind conditions of a certain area?

Wind speed is constantly changing as the Earth's atmosphere is a highly non-linear, even chaotic system that can be described only approximately [44]. It is therefore necessary to use statistical tools to describe the wind; furthermore, the statistical distribution of wind speed is of great interest, because it directly reveals whether a certain place has enough wind to be economically interesting for building wind power. Normally, one uses the Weibull distribution to describe the wind speed frequency [64]. Other possibilities include the gamma, lognormal, three parameter beta and Rayleigh distributions [2]. The *Probability Density Function* (PDF) of Weibull distribution for wind speed w is written as:

$$f_{WB}(w) = \frac{\nu}{\varsigma}(\frac{w}{\varsigma})^{\nu-1}e^{-(\frac{w}{\varsigma})^{\nu}} \tag{14}$$

The parameters (ν, ς) of Weibull distribution have to be estimated from historical wind speed data. The normal estimation procedure is the maximization of the likehood of parameters, but other methods have been proposed as well [2]. The distribution characteristics of wind speed may however change in time due to change in local vegetation, *roughness* and so on. This is why care must be taken in order to update the information regarding the distribution of wind speed.

Recently, the Finnish Meteorological Institute (FMI) published a new Wind Atlas [36] which summarizes the prevailing wind conditions throughout Finland by using a statistical distribution of the wind speed. The simulations are carried out using an ensemble of *Numerical Weather Prediction* (NWP) models. Those kinds of models will be further described later in this thesis (Section 4.1). The same kind of wind atlases are used in other European countries as well [28].

Owing to the fact that often wind speed can be characterized using a Weibull distribution, one clearly sees that relatively low wind speeds are more common than

steady winds. Looking at the power curve (Figure 1), this means that normally the power produced can be seen coming from the left and center parts of the curve. Specifically in the center part of the curve a small change in wind speed leads to a large change in power. This is one reason why wind energy is often called *intermittent* in literature. Figure 3 gives a graphical illustration on this.

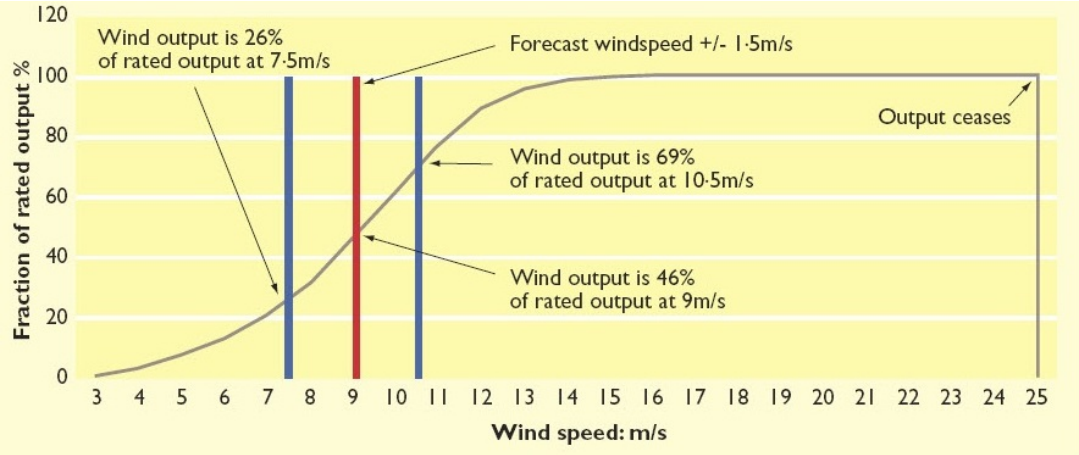

Figure 3: Illustration of the effect of wind speed change on the generated power. Source: wesrch1.wesrch.com

However, wind turbines do not normally operate at their full capacity simply because it would mean the locations of wind turbines would be experiencing very strong winds all the time. For measuring the ratio of true power output to the theoretical rated power a *capacity factor* is introduced. Nowadays the on-shore wind farms have a capacity factor of 25%–40%, while in off-shore parks this can go up to 45%–60% [64].

Managing the intermittence of wind power generation is a key issue in the integration of wind energy into electricity grids [32, 81]. As stated before, all the conventional power generation methods are *dispatchable*, so that their level of production is always directly imposed by operators. This is not the case with wind power; therefore it is a true challenge for electricity grids, more specifically the *Transmission System Operators* (TSOs), whose job is to maintain the balance between electricity supply and demand continuously. Actually wind power can even be seen as a *negative load* [64], which must be tracked as accurately as possible.

Generally speaking, the electricity producers must schedule their production in advance in order to respond to electricity demands and its variations. This is quite easily done with conventional power generation, owing to its ability to be dispatched. Production scheduling is based on modeling and forecasting electricity load in a particular area. Such modeling has gained much attention in the previous years in literature, and thus the mean absolute errors related to load forecasting are nowadays approximately 1.5%–2% for a 24 hours forecast [4, 86].

Because wind power is intermittent, it is necessary to be prepared for unexpected drops in wind power production. Examples of such drops are for instance severe

storms, when wind turbines are required to be switched off in order to prevent equipment damage; an unpredicted decrease of wind speed, or a breakage of many wind turbines at the same time. This means that the TSOs have to quantify the need for reserve power in order to be able to compensate such sudden losses of wind production. Another possibility is to use energy storage devices, which permit to store wind power generation. However, today such devices are quite expensive and not economically viable in the near future [8]. There has been research concerning the use of hydro-storages, which would allow wind and hydro electricity producers to generate an even larger revenue acting conjunctively, vice separately. An example of this kind of investigation can be found at [53, 5].

Summarizing the discussion so far, it seems that especially when the wind power penetration is high, i.e. when the system uses a lot of electricity produced by wind, the problems related to the intermittence are also notorious, inevitable and imminent. There exists, however, another side of the coin: as fluctuations of wind power output are greater using a single turbine than in a farm consisting of several turbines, this is also true when wind production is distributed over a large geographical area having different wind characteristics. This effect is known as the spatial smoothing and is studied further in [32].

There also exist mathematical tools and mechanisms for managing such omnipresent and never fully cancellable intermittency of wind power production. Even though one cannot fully compensate for the chaotic nature of wind and is thus subject to variations of the scheduled production, by addressing the production predictions with an estimation on their uncertainty, wind production becomes more dispatchable in statistical terms.

3 Electricity markets

This Section provides an overview on the general structure of markets, where electricity is bought and sold. This helps us to understand the challenges a wind power producer encounters when it enters an electricity market. With the new legislation related to wind power subsidies, producers are also actually being forced [60] to sell the electricity produced by its resources in such markets. Furthermore, we insist on the role of balancing supply and demand on a continuous basis, which is the responsibility of a local transmission system operator. We also discuss the penalties due to the imbalances of this non-dispatchable means of production, and investigate possibilities for the wind power producer to better manage its production means.

Before going further, it must be noted that in addition to physical electricity markets, a significant volume of electricity is traded with several types of financial contracts. In this thesis we nevertheless omit the treatment of these financial electricity markets as the nature of wind power is not very suitable for trading on such floors. There is however research concerning the use of weather-related options in hedging the wind power variability [31]. A review on financial electricity markets can be found in [84].

3.1 Electricity wholesale markets

In the last two decades the electricity industry has experienced significant changes due to the effects of deregulation and open market competition. The aim of this evolution has been the improvement of economic efficiency, hence the hope for electricity price reduction [38]. Although electricity is still commonly bought and sold *Over-The-Counter*, in which the buyer and seller directly make a deal concerning the price and volume of electricity, those changes have in many places culminated in the appearance of a wholesale electricity market. Examples of wholesale markets are the Nordic NordPool and the French/German EPEX.

The evolution towards greater unification and deregulation is far from over; for instance in the Nordic market there are plans for adding the Baltic countries to the common market area [25]. In this new context, the actual operation of the electricity generating units no longer depends on the state- or utility-based centralized procedures, but rather on decentralized decisions of generation firms [85], whose goals are purely economic: they aim at maximizing their own profits. All firms compete to provide generation services at a price set by the market, which itself is a result of the interaction of the generation firms and the demand for electricity.

Two main features characterize electricity as a commodity: it has a very limited, if not zero storability and its transportation requires a physical link in order for it to transpire. The inability to store electricity makes electricity produced and delivered at different times distinct commodities. For instance if a buyer is willing to buy x MWh at a specific moment, they normally cannot be satisfied in having it 6 hours or days later. Generally speaking this fact makes electricity prices strongly dependent on electricity demand, although large deviations still occur due to different behaviors

of electricity buyers and suppliers [40, 38].

The limitations of transportation mean that between certain regions, the transmission of electricity may not be economically feasible, or possible. The prices of electricity are thus highly localized, reflecting the local determinants of supply, such as plant characteristics and weather, and demand [51], for instance when water heaters are all switched on at the same time.

3.1.1 Spot markets

The most active marketplace for electricity wholesale is a day-ahead or spot market, of which the NordPool Elspot is one of the oldest, and as such, a good example. The actors in the spot market, i.e. the producers and distributors (buyers) of electricity, express their willingness to sell and buy electricity at a certain price and volume. This procedure is called an *equilibrium point trading*, and is carried out once a day for every hour of the next day. The price for tomorrow is thus decided today, usually around noon [70]. This is where the electricity market itself takes its active role: it gathers up all bids and asks sent to it, and aggregates them into inverse supply and demand curves. These curves represent the collective, aggregate wishes of the actors in the market.

How the single bids and asks are formed is out of the scope of this thesis. We simply express that normally the suppliers of electricity price their production according to their marginal costs, which is in accordance with the classical microeconomic theory [83]. It can be already noted here that for the wind power producers these marginal costs of production are zero. Respectively, the buyers normally simply want to buy electricity at a minimum price, respecting their estimates of its demand for the next day.

As stated above, the transmission of electricity from a place where it is generated to another place where it would be consumed can be limited, if not totally impossible. Because of this it is quite normal that the day-ahead markets for electricity are divided into price areas, according to the geographical limits imposed by difficulty of electricity transmission. In the case of the Nordic market, there are several price areas depending on *bottlenecks* of transmission. For further description of price areas see [59].

Let us now introduce the corresponding inverse supply $S_h^a(x)$ and demand $D_h^a(x)$ functions for every area a and hour h. They both put out a price given a volume argument. Then we define the *social surplus* as the substraction between the *consumer surplus* and the *producer surplus*, which are calculated up to the equilibrium supply s_a^h and demand d_a^h [83]. The maximization of this function solves the area prices for every individual hour.

$$\max \sum_a \left\{ \int_0^{d_a^h} D_h^a(x)dx - \int_0^{s_a^h} S_h^a(x)dx \right\} \tag{15}$$

The maximization procedure of (15) is nevertheless subject to various constraints due to the very nature of electricity as a commodity. They can be listed as follows:

- All the bids and asks are between their upper and lower limits
- Concerning bids and asks, every area is in balance at every hour, which means that supply + import - export equals demand
- There are physical limits for the transmission between different areas

It can be shown that the maximization problem (15) with the constraints above yields a unique solution, as long as the functions $S(x)$ and $D(x)$ are monotone [59]. The Lagrangian dual variable [13] of the problem (15) can be interpreted as the spot price corresponding to each area and hour.

3.1.2 Intraday markets

A power company may encounter imbalances between its realized production and the one fixed at the day-ahead market. The reasons for these imbalances are various: power units may break down or generated power does not match the amount planned. Especially the latter reason is very common among the wind power producers (WPPs). Generally, all the producers want to correct their predicted imbalances before it is too late, i.e. before they really have to deliver the amount of electricity they have promised to do in the day-ahead market. Consequently, there normally exists a market mechanism for making adjustments to these day-ahead trades. This kind of *intraday market* reduces the power companies' risk of being in imbalance, because the price for electricity traded in the intraday market is known prior to the hour of delivery rather than afterwards [59]. In Section 3.2 we show that the highly volatile prices of the actual balancing market, which will be known only after the time of delivery, carry a rather large economical risk.

An example of this kind of intraday market is the Nordic Elbas. It provides a platform for making trades, which can happen within and across the price areas included in the Elspot market. In the Elbas, trading is possible around the clock until one hour before the physical delivery of electricity.

As there is a large set of different actors with a substantial spread in their production mixes and demand estimates, Elbas provides an opportunity to make profitable deals for both of the market counterparts. Producers reduce their risks and get a better control of their production means, and consumers increase their chances of procuring electricity for correcting deviations which occur after the day-ahead estimates [59]. Generally the increased liquidity of the markets is for the benefit of both.

The price forming mechanism is somewhat different from the spot market. There is no single equilibrium price for all of the market actors. On the contrary, the market works like a stock exchange. Buyers and sellers of electricity search for a

volume corresponding to their needs and pick it up if its price announced by the counterparty is acceptable.

There is however empirical evidence that the liquidity of intraday markets can be rather low, while the volumes traded in the Elbas market are sometimes rather limited and few actors use it actively. Thus the possible benefits, which are contingent to market liquidity, may be volatile. In addition, without the presence of abnormal demand or supply patterns, the Elbas market prices generally resemble the Elspot prices. For illustration purposes, a randomly chosen 100 hour period of Finnish Elspot prices, average Elbas prices as well as Finnish electricity consumption are presented in Figure 4.

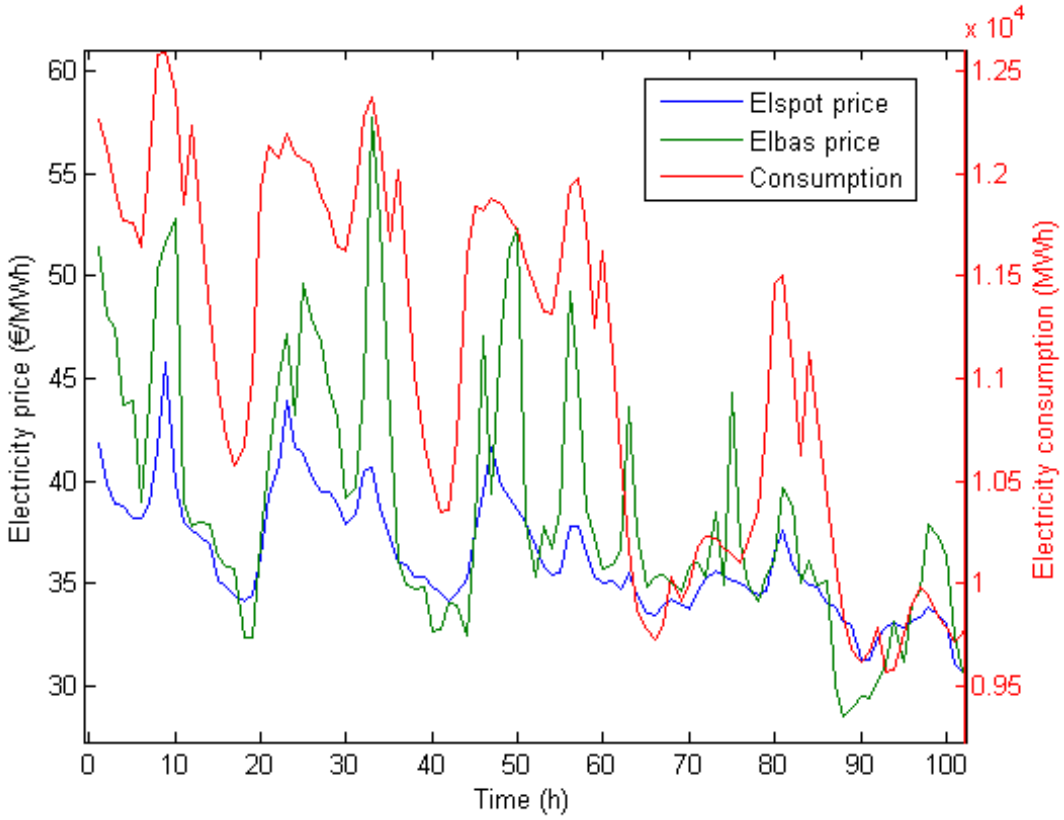

Figure 4: Finnish electricity consumption and spot & intra-day prices in a 100 hour period of the year 2009. Data source: [59]

Summarizing the discussion above, the revenue obtained by an electricity generating company participating in an electricity market can generally be written as:

$$R_t(\pi_{d,t}, \pi_{i,t}) = \lambda_{d,t}\pi_{d,t} + \lambda_{i,t}(\pi_{i,t} - \pi_{d,t}) + IC_t = r(\underbrace{\cdot}_{\text{RV}}; \underbrace{\cdot}_{\text{DV}}; \underbrace{\cdot}_{\text{PAR}}) \qquad (16)$$

with R_t denoting the net revenue, $\pi_{d,t}$ the power committed to day-ahead markets, $\pi_{i,t}$ the power committed to intraday markets, $\lambda_{d,t}$ and $\lambda_{i,t}$ the respecting prices at

those markets and IC_t the imbalance cost. Such formulation of the power producer revenue can be found for example in [80]. The revenue may thus be decomposed into (maximally) three components corresponding to random variables (RV), decision variables (DV) and parameters (PAR). The purpose of this kind of formulation will become clear in the subsequent Section. We next discuss the formulation and characteristics of the imbalance cost IC_t.

3.2 Balancing supply and demand

As electricity cannot be stored in any reasonable way, an electric power system must be continuously balanced. This is the prime responsibility of national TSOs who manage the national balance. They oblige actors to carry out methods for evaluating production and consumption in their balances. Every single market player must have a *balance responsible party* which in turn has a responsibility to plan its production and consumption in balance. If imbalances occur, they are actually seen only after the hour in question, when a balance reporting is made. However, what is essential here is that every imbalance must be annulled. Such operation incurs either additional costs (one has to buy more energy) or revenues (one sells the surplus to the markets). However, we will below see that these cash flows can always be seen as punishing when compared to a situation of perfect predictions.

Actually the Finnish balancing sheet is divided into two parts, emphasizing the two sides of electricity use: production and consumption. This division is described in Figure 5. However as this thesis mainly aims at characterizing the implications of producer behavior, we will hereafter concentrate on the production balance described on the left part of Figure 5. Generally a WPP has to maintain the consumption side of the balance sheet as well, because the wind power turbines require a certain amount of energy in keeping them operationally available. Such an effect is however negligible and will not be discussed further.

3.2.1 Regulatory market

In order to handle any unpredictable differences between the planned and the real exchange during delivery once the day-ahead and intraday markets are closed, the national transmission system operators have additionally set up a regulating market from which required upward or downward regulation is obtained on a short notice [51]. In these markets, every power producer or large user can make bids according to its specific ability and willingness to gain from helping others in reducing their imbalances. The bids are grouped in two categories corresponding to the willingness to increase (*up*) and decrease (*down*) production, respectively; then the national TSO activates them if there is a need to do so.

As the nature of wind power is not suitable (it cannot up-regulate at all, and down-regulation possibilities are limited) for taking such market actions on a short notice, a WPP cannot take part in the regulating markets. Consequently, a WPP has

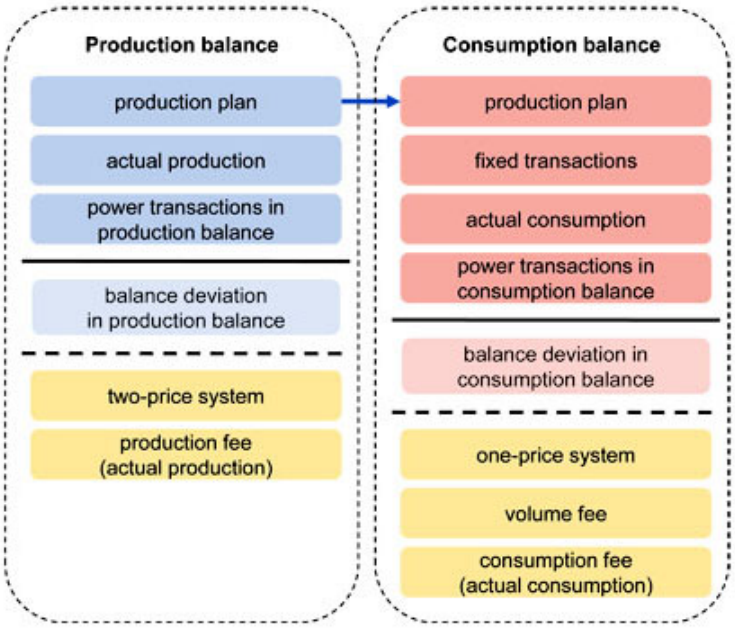

Figure 5: Finnish balancing sheet. Source: http://www.fingrid.fi

nothing to gain in the form of making profit from others' imbalances; it can only lose in the form of being forced to pay for its own deviations. It is thus relevant to see WPP as a price taker in what it comes to balancing. In this sense, the balancing responsibility can be seen as a sort of punishment for the WPP imposed by its inability to schedule and forecast its production properly.

The functioning of regulatory markets can be described by assuming a rational behaviour of the actors participating in it. The producers are willing to down-regulate their resources at a smaller price than their marginal production costs. Keeping in mind how the spot price is formed, we have a down-regulation price $\lambda_{down,t} < \lambda_{d,t}$. Respectively, the rational behaviour implies that producers are willing to produce additional energy required to cover the deficit at an up-regulation price higher than the spot price. Thus for the up-regulation price we have $\lambda_{up,t} > \lambda_{d,t}$.

3.2.2 Balancing cost

There are several different mechanisms concerning the price formation of balancing energy, but here we describe the Finnish model in depth. The responsible counterpart of the balancing market is the Finnish TSO Fingrid. The balancing market follows a *dual imbalance pricing* [62], where a different price is applied to positive (*long*) and negative (*short*) imbalance volumes of production. Furthermore, in the Finnish system, the balancing responsibility is a function of the balance of the whole transmission system: if a power generator has its imbalance in the opposite direction than the whole system hence helping the system to cope with its imbalance, there is no extra cost for its imbalance. This type of pricing is described in Figure

6, where two successive similar colours represent a balancing cost, and conversely two different colours an exemption from balancing penalty. It is motivated by the discussion in [65, 56, 48].

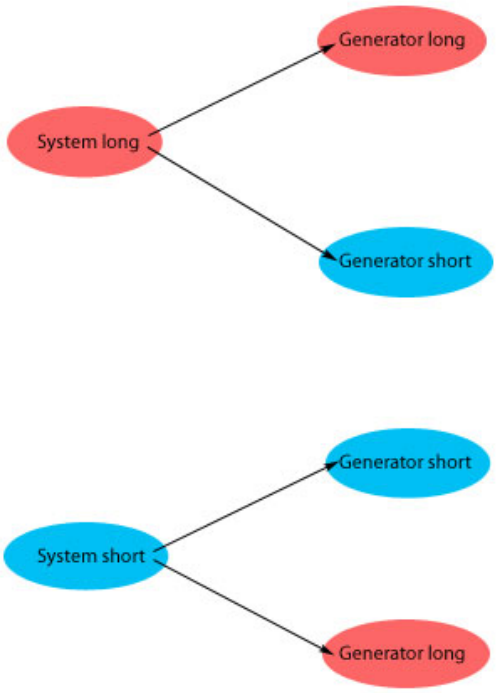

Figure 6: Finnish electricity balancing prices as a function of imbalance signs: two successive similar colours represent imbalance cost, two different colours an exemption from it

We now define a positive energy deviation to be a higher production than scheduled, and a negative deviation respectively to be a lower production than scheduled. The prices $(\lambda_{+,t}, \lambda_{-,t})$ for these imbalances representing the cost of the energy required to offset the unplanned deviations in the Finnish balancing framework can be mathematically written as follows:

If the system is long (has a positive imbalance, an excess of power), then

$$\begin{cases} \lambda_{+,t} = \min(\lambda_{d,t}, \lambda_{down,t}), & \text{if generator long} \\ \lambda_{-,t} = \lambda_{d,t}, & \text{if generator short} \end{cases} \tag{17}$$

This means simply that the market participants causing the system to be long are remunerated for their overproduction at a price $\lambda_{+,t}$. They obtain a smaller profit than they would have if their production surplus had been sold in the spot market. Those who are short, i.e. producing less energy than scheduled and helping the system to cope with its total imbalance, are remunerated at the spot price for their

realized production. This means that being on the opposite side as the system does not incur an additional balancing cost.

If the system is short (has a negative imbalance, a deficit of power), then

$$\begin{cases} \lambda_{+,t} = \lambda_{d,t}, & \text{if generator long} \\ \lambda_{-,t} = \max(\lambda_{d,t}, \lambda_{up,t}), & \text{if generator short} \end{cases} \tag{18}$$

As the price of additional energy to cover the deficit is higher than the one at the spot market, the producers incur a profit loss compared to the situation where they would have been apt to sell the correct amount in the day-ahead market. As previously, those on the other side helping the system will be paid at the spot price for their actual production.

Note that looking at (17) and (18) one can see that clearly $\lambda_{+,t} \leq \lambda_{d,t}$ and $\lambda_{-,t} \geq \lambda_{d,t}$. We can also calculate ratios between the observed balancing and spot prices as follows:

$$\begin{aligned} r_{+,t} &= \frac{\lambda_{+,t}}{\lambda_{d,t}} \\ r_{-,t} &= \frac{\lambda_{-,t}}{\lambda_{d,t}} \end{aligned} \tag{19}$$

A schematic illustration of Finnish balancing price ratios over a clip (26.1.2009 – 5.2.2009, note that there are some missing data points) from original data (see Section 6.1) is presented in Figure 7.

It must again be emphasized that those market participants who contribute to the system imbalance are punished and those helping the system are not. This seems quite reasonable and can be seen as apologizing unintentional deviations and penalizing purposeful ones.

Now bearing in mind the above discussion on the regulatory and imbalance prices, the actual imbalance cost introduced in Section 3.1.2 can be formulated. If the realized production at time t is $\pi_{g,t}$, then the imbalance cost is written as

$$IC_t = \begin{cases} \lambda_{+,t}(\pi_{g,t} - \pi_{i,t}) & \pi_{g,t} > \pi_{i,t} \\ \lambda_{-,t}(\pi_{g,t} - \pi_{i,t}) & \pi_{g,t} < \pi_{i,t} \end{cases} \tag{20}$$

3.3 Wind power producer in electricity markets

What applies to generators possessing conventional power generating units applies to wind power producers as well: they all want to maximize their profit from selling their production to electricity markets. At this point there is no distinction between the behaviors of power generating units irrespective of what their production means are. However, a WPP encounters an additional challenge compared to other

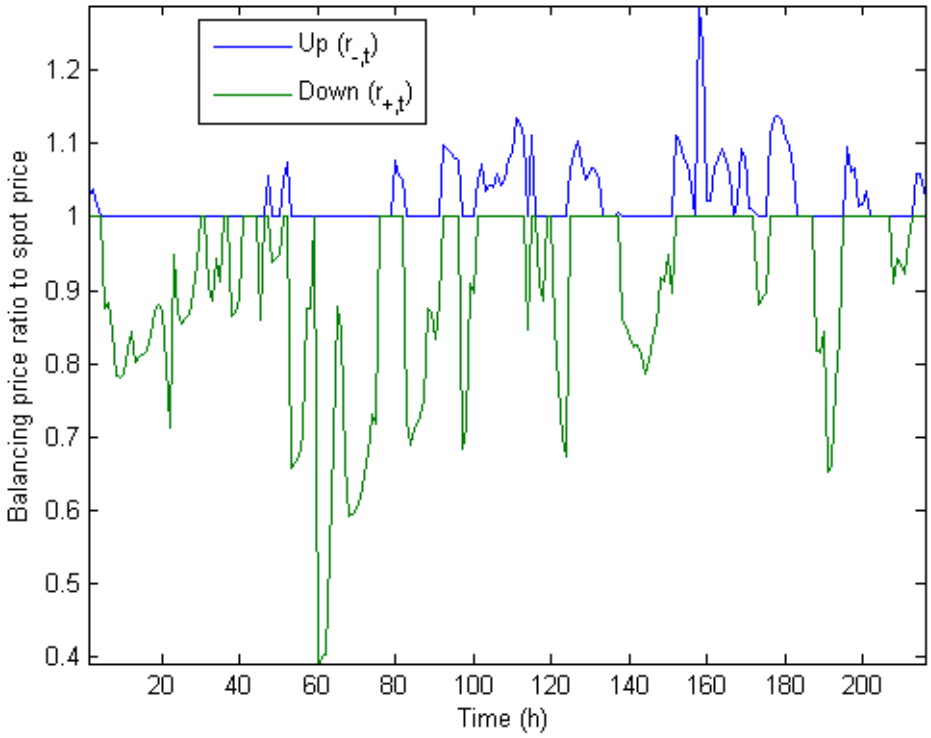

Figure 7: An illustration of Finnish balancing prices 26.1.2009 – 5.2.2009

conventional production forms as it cannot know exactly how much power it will generate in the future. If one looks at conventional power (e.g. thermal, hydro or nuclear) producers, it will be noticed that their command to their resources is complete, without taking into account possible disturbances. A WPP can only estimate its production and is thus more vulnerable to deviations from it, which will be penalized, as seen in Section 3.2.

The chronological typical behavioral pattern of a WPP can be summarized as follows. To begin with, in order for it to be possible to predict the production with a horizon of 13–36 hours ahead (remember that the gate closure for day-ahead markets is usually around noon), a WPP must possess a numerical weather prediction of future winds corresponding to those future instants of the next day. Then, using a power prediction model, it forms the predictions for its production for the next day and submits the obtained volumes to the electricity day-ahead market at a marginal price. This figure is usually zero, but can be marginally above zero as well as there may be some volume-based charges imposed by the TSO when feeding electricity into the grid. In some markets even negative prices are possible, but what comes to wind production, without a fear of imbalances it will always be profitable to sell the generated electricity at any positive price.

After the day-ahead market has closed, the prices for next day's electricity for every hour are obtained. At this point, the revenue of a WPP is a simple multiplication of

the predicted energy and the now certain spot price, so long as no deviations from the fixed production occur. While such almost coincidental fact rarely happens, the WPP will then consider actions which can improve its position before the imbalances are charged. A possibility for these corrective actions is participation in intraday markets (see Section 3.1.2).

As will be shown later, it will be profitable to run the power predictions again for shorter horizons in order to decrease the inherent uncertainty of wind production. If by running these estimations at successive instants, a WPP estimates its production for a specific hour cannot fulfill the production level it originally scheduled, and has promised to deliver, it can try to buy or sell the estimated lack or surplus in an intraday market. For this purpose it looks at the corresponding offers to buy and sell and chooses an appropriate amount of energy at a price it deems acceptable.

In the next Section, different trading strategies are elaborated based on the need of designing efficient methods to cope with the profit loss imposed by uncertain predictions. Before going any further, we wish to note to the reader that all of the following strategies are based on some fundamental assumptions, of which the most important is that a single WPP has no market power. This means that its bidding decisions do not directly affect the equilibrium price in the day-ahead, intraday nor the corrective markets. We ask the reader to refer to e.g. [76] if one wishes to consider such effects. Furthermore, the marginal bidding price is assumed to be zero, so that if there are no negative prices the WPP bids are always accepted.

3.4 Trading under uncertainty

Sections 4 and 6 concentrate on providing accurate forecasts of future wind production. There is however an inherent uncertainty that constitutes wind availability and thus wind power production. Although this uncertainty diminishes as the forecasting horizon decreases, it cannot ever fully be cancelled. It means that a WPP is always to some extent plagued by this uncertainty, which in its turn deteriorates its profits gained from selling the produced energy to the markets. This Subsection brings particular focus to the role of analyzing the uncertainty and furthermore how such a knowledge can be harnessed in order to dampen the profit variability by producing optimal bidding strategies.

We denote $\pi_{g,t}$ the realization of a random variable $\Pi_{g,t}$, which describes the possible outcomes of future power production at instant t. Now, the forecast provided by any kind of *point forecasting system* can be (normally) described as being the conditional expectation of $\Pi_{g,t}$, based on the knowledge of the given model f, its parameters $\aleph$ and its inputs Φ_τ [64]. Mathematically it may be written as $\hat{p}_{t|\tau} = E\left[\Pi_{g,t}|f;\aleph;\Phi_\tau\right]$ [67]. The forecast on which the bids traditionally are based hence minimizes the sum of squared errors between the forecasting model fit and the realized production. By nature it is only a summary statistic, a single number; therefore it cannot tell what *could* happen. As will be seen, harnessing such knowledge can increase the WPP profits.

At first, we assume that a WPP wishes not to trade at any kind of intraday (corrective) markets, but rather makes its initial power scheduling in the spot market and then accepts the costs incurred by possible deviations from the predicted production. It can be shown [65, 48] that the prediction-based bid it makes on the day-ahead market may not be the optimal one it should bid in order to hedge against the balancing costs. This is done by investigating the formula for the WPP profit (see (16), (20)):

The revenue for a bid $\pi_{d,t}$ is

$$R_t = \lambda_{d,t}\pi_{d,t} + \begin{cases} \lambda_{+,t}(\pi_{g,t} - \pi_{d,t}), & \pi_{g,t} > \pi_{d,t} \\ \lambda_{-,t}(\pi_{g,t} - \pi_{d,t}), & \pi_{g,t} < \pi_{d,t} \end{cases} \tag{21}$$

By adding and substracting $\lambda_{d,t}\pi_{g,t}$ on both sides and collecting appropriate terms (21) may be written as

$$R_t = \lambda_{d,t}\pi_{g,t} - \begin{cases} \overbrace{(\lambda_{d,t} - \lambda_{+,t})}^{\geq 0}\overbrace{(\pi_{g,t} - \pi_{d,t})}^{>0}, & \pi_{g,t} > \pi_{d,t} \\ \underbrace{-(\lambda_{-,t} - \lambda_{d,t})}_{\leq 0}\underbrace{(\pi_{g,t} - \pi_{d,t})}_{<0}, & \pi_{g,t} < \pi_{d,t} \end{cases} \tag{22}$$

The first term in (22) corresponds to the revenue which one obtains by selling the actually produced energy at the spot price. The second term corresponds to the profit loss due to imperfect forecasting. Note that both multiplied terms in the curly brackets are non-negative, so that if producer is not mitigating the system imbalance, every deviation from the perfect prediction yields a profit loss. The initial scheduling may now be searched in a pursuit to maximize the total profit. As one cannot control $\lambda_{d,t}$ or $\pi_{g,t}$, such procedure itself is equivalent to searching for a bid $\pi_{d,t}$ that minimizes the balancing cost (i.e. the second term in (22)).

We define the deviation between the realized and scheduled production as δ_t, which is a realization of its corresponding random variable. The balancing cost can be written as a function of the realized deviation:

$$g(\delta_t) = \begin{cases} (\lambda_{d,t} - \lambda_{+,t})(\pi_{g,t} - \pi_{d,t}) = (\lambda_{d,t} - \lambda_{+,t})\delta_t, & \delta_t > 0 \\ -(\lambda_{-,t} - \lambda_{d,t})(\pi_{g,t} - \pi_{d,t}) = -(\lambda_{-,t} - \lambda_{d,t})\delta_t, & \delta_t < 0 \end{cases} \tag{23}$$

The most straightforward way in obtaining an optimal bidding strategy is to be risk-neutral, so that there is no preference towards valuing being long differently than being short. Then we can consider the *expectation* of the balancing cost defined in (23). Assuming independence of generated power and electricity prices (no WPP market power), the minimization of the excepted balancing cost can be formulated

as

$$\begin{aligned}\min_{\pi_{d,t}} E\left[g(\delta_t)\right] = &- E\left[(\lambda_{-,t} - \lambda_{d,t})\right] \int_0^{\pi_{d,t}} (\pi_{g,t} - \pi_{d,t}) f_\Pi(\pi_{g,t}) d\pi_{g,t} + \\ &E\left[(\lambda_{d,t} - \lambda_{+,t})\right] \int_{\pi_{d,t}}^1 (\pi_{g,t} - \pi_{d,t}) f_\Pi(\pi_{g,t}) d\pi_{g,t}\end{aligned} \tag{24}$$

Keeping in mind the definitions (17), (18) for imbalance prices, we obtain the expectations of the unit regulation costs $(\lambda_{-,t} - \lambda_{d,t})$ and $(\lambda_{d,t} - \lambda_{+,t})$:

$$\begin{aligned}E\left[(\lambda_{-,t} - \lambda_{d,t})\right] &= \omega_{-,t}\lambda_{up,t} + (1 - \omega_{-,t})\lambda_{d,t} - \lambda_{d,t} = \omega_{-,t}(\lambda_{up,t} - \lambda_{d,t}) \\ E\left[(\lambda_{d,t} - \lambda_{+,t})\right] &= \lambda_{d,t} - (1 - \omega_{+,t})\lambda_{d,t} - \omega_{+,t}\lambda_{down,t} = \omega_{+,t}(\lambda_{d,t} - \lambda_{down,t})\end{aligned} \tag{25}$$

where the terms $\omega_{+/-,t}$ represent the probability of the system needing downward or upward regulation, respectively.

The key factor in (24) is the term $f_\Pi(\pi_{g,t})$, which describes the *probability density* of producing $\pi_{g,t}$ in the future. Let us now introduce a notation change of the form

$$\begin{aligned}\tilde{\lambda}_{+,t} &= \lambda_{d,t} - \lambda_{down,t} \\ \tilde{\lambda}_{-,t} &= \lambda_{up,t} - \lambda_{d,t}\end{aligned} \tag{26}$$

so that we have

$$\begin{aligned}\min_{\pi_{d,t}} E\left[g(\delta_t)\right] = &- \omega_{-,t}\tilde{\lambda}_{-,t} \int_0^{\pi_{d,t}} (\pi_{g,t} - \pi_{d,t}) f_\Pi(\pi_{g,t}) d\pi_{g,t} + \\ &\omega_{+,t}\tilde{\lambda}_{+,t} \int_{\pi_{d,t}}^1 (\pi_{g,t} - \pi_{d,t}) f_\Pi(\pi_{g,t}) d\pi_{g,t}\end{aligned} \tag{27}$$

Taking the derivative of (27) with respect to the bid $\pi_{d,t}$ and by applying the Leibniz rule we obtain the necessary condition for the optimum:

$$\frac{\partial E\left[g(\delta_t)\right]}{\partial \pi_{d,t}} = (\omega_{+,t}\tilde{\lambda}_{+,t} + \omega_{-,t}\tilde{\lambda}_{-,t}) F_\Pi(\pi_{d,t}) - \omega_{+,t}\tilde{\lambda}_{+,t} = 0 \tag{28}$$

Here F denotes the *Cumulative Distribution Function* (CDF) of $\pi_{g,t}$. Solving (28) (the production CDF F is monotonically non-decreasing) we obtain the unique solution for the optimal bid:

$$\pi^*_{d,t} = \arg\min_{\pi_{d,t}} E\left[g(\delta_t)\right] = F_\Pi^{-1}\left(\frac{\omega_{+,t}\tilde{\lambda}_{+,t}}{\omega_{+,t}\tilde{\lambda}_{+,t} + \omega_{-,t}\tilde{\lambda}_{-,t}}\right) \tag{29}$$

The term $F_{\Pi}^{-1}(\cdot)$ in (29) corresponds to a *quantile function*, which is the inverse of the cumulative distribution function of possible future production values. It may be mathematically defined as

$$F^{-1}(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\} \tag{30}$$

It is thus shown that without considering any corrective actions, the optimal bid is a certain quantile of the future production probability distribution and *not* the value one would obtain by using a point forecasting system, whatever its type might be. The quantile is uniquely determined by the relative balancing prices and the probabilities of the system state corresponding to the future instant.

There is an intriguing property of (29) worth mentioning: should the WPP know exactly (beforehand) the relative balancing prices $\tilde{\lambda}_{+/-,t}$, the quotient communicating the optimal quantile in (29) would always yield either zero or one. This means that if the system needs electricity, hence activating up-regulation, the term $\tilde{\lambda}_{+,t} = 0$, which entails that the optimal quantile is zero. Consequently the WPP should bid a production which falls under its initial power prediction estimate with a zero probability; the only possibility for that would be bidding zero production. With such a bid, the WPP would then gain from its any realized production in the balancing markets at the spot price. The same kind of reasoning is valid if the system has too much electricity and hence activating downregulation; in that case $\tilde{\lambda}_{-,t} = 0 \Rightarrow \pi_{d,t} = 1$, so that it would be optimal for the WPP to bid a full production, which is again equivalent to gaining spot revenue from the realized production. In sum, perfect predictions of imbalance prices are exactly as valuable as perfect power predictions from a strategic point of view [56], provided that regulation is undertaken by the TSO. However such 'strategic' bidding activities cannot be seen as very stabilizing from the point of view of the total system. Furthermore, following the Finnish subsidy legislation only electricity sold in the markets is recompensed at the subsidized price, so that zero bidding is not rational.

If a WPP rather positions itself as being risk-adverse (wanting to minimize large economical losses) the risk-neutrality assumption used above might not be relevant. One solution has been presented in [65], where minimization of the worst possible scenario incurring losses leads to a day-ahead bid defined as

$$\pi_{d,t}^{*} = \arg\min_{\pi_{d,t}} \max_{\pi_{g,t}} g(\delta_t) f_{\Pi}(\pi_{g,t}) \tag{31}$$

Another possibility is to consider the spot trading as usual, with respect to best available point prediction regarding every future hour. Then with a fairly similar approach as above, it can be calculated what is the optimal position to take in intraday markets *given* the day-ahead position in order to dampen the balancing cost variations. We will present such formulation briefly, following the discussion in [80].

The starting point is again (16) which describes the revenue WPP attains by selling

the production both to spot and intraday markets. Now, assuming that the day-ahead bid $\pi_{d,t}$ is not a subject of optimization, but rather is decided based on forecasting information on future production, we can optimize the intraday bid $\pi_{i,t}$ with the aim of maximizing the total profit. If the realized production $\pi_{g,t}$ and the intraday price $\lambda_{i,t}$ are considered as independent random variables, we have

$$\max_{\pi_{i,t}} E\left[R_t\right] = \iint_{-\infty}^{\infty} r(\pi_{g,t}, \lambda_{i,t}; \pi_{i,t}; \pi_{d,t}, \lambda_{d,t}, \lambda_{+,t}, \lambda_{-,t}) f_{\Pi}(\pi_{g,t}) f_{\Lambda}(\lambda_{i,t}) d\pi_{g,t} d\lambda_{i,t} \quad (32)$$

Such maximization program must be solved independently for every instant t. Solving the program may be easier if one first discretizes (32).

We can also consider a situation where a WPP wants to optimize *both* the bids in day-ahead and intraday markets. In such a situation, we cannot simply formulate a revenue maximization (or a balancing cost minimization as in (24)) as it can be too difficult to express the probability densities related to spot price, intraday price, balancing price, wind power generation and the number of available turbines. Furthermore it can be argued that for instance the spot and intraday prices are definitely not independent of each other, so that a simple multiplication of probability densities carried out in (32) would be out of question. Another reason for the increased difficulty in forming an analytic solution is that as the uncertainty of wind power generation decreases as the time of physical delivery gets closer, the interdependence structure of forecast errors among different look-ahead times is neglected [66]. Harnessing the knowledge of such a structure would however be essential in considering participating in multiple markets.

Summarizing the reasoning above, if a WPP wants to maximize its profits from the day-ahead and intraday markets and simultaneously minimize the expected balancing cost, accounting for the different uncertainties described above, it must characterize the uncertainties as stochastic processes and solve the optimizing problem by using a *multi-stage stochastic programming* approach. In such a programming formulation, each stage represents a successive market wherein decisions are made based on scenarios formed by knowledge of the relevant variables so far. Examples of this kind of modeling are published in [56, 54]. Without dwelling into details, [56] claims that using a stochastic programming approach the *value of stochastic solution* [10] can go up to approximately 3.3%.

Up to now, the art of obtaining the probability density (or cumulative distribution) function of the random variable $\Pi_{g,t}$ has been neglected. This issue is discussed in Section 4.3.5, which presents several literature-based approaches in finding a probability distribution for the future production.

At last, it must be acknowledged that in (24) and (32) the balancing prices $\lambda_{+/-,t}$ are assumed to be known at every future instant t. In reality however, they are not known at the time of making bids for spot nor intraday markets, but have to be predicted. Therefore the obtained bid will naturally be optimal only if the prediction of balancing prices is correct [48]. There are methods for constructing such price

scenarios (and thus forecasting them to some extent, see e.g. [61]). In any case it may turn out to be quite difficult to forecast the balancing prices accurately (and furthermore to take into account the fact that the system total balance influences the balancing responsibility).

It is nevertheless shown that the price terms in (27) can be replaced by their expectations [48]. The expectations can then be estimated by historical observed values. In the literature tests have been made using averaged balancing prices over a predefined time window. The results (see e.g. [65]) normally show an increase of profit compared to the situation in which no strategic bidding is performed. The increase is nevertheless smaller than it would be using perfect price predictions. This gap is called the *expected value of perfect information* [10].

4 Wind power forecasting methods

In the previous Section, as in numerous other studies (see e.g. [7]), it has been shown that the economic benefits of good and reliable forecasting are substantial. We can now turn our attention towards the art of forecasting wind generation. The questions one might ask are the following: How can these benefits be attained? What is a good forecaster like? What types of forecasting methods are available?

This Section is divided into three parts. First we describe basic methods for forecasting wind speed using numerical weather models. Then in the second part we discuss the general principles related to forecasting power generated by wind, and present some measures used in assessing the performance of a forecaster. The third and final part peruses through several wind power forecasting models found in the literature.

4.1 Numerical weather prediction

The weather conditions in an area where wind power is located are the most important factors linked to wind power generation. It is thus absolutely necessary to predict the appropriate weather variables in order to have a prediction on wind power. This weather prediction process based on *meteorology* is not the subject of this thesis; however in order to obtain a better understanding of the behavior of wind power forecasts later, the numerical weather prediction is briefly discussed here.

The meteorological weather prediction is based on numerical weather prediction models. They are basically a large set of highly non-linear intercoupled partial differential equations describing the forces and motions related to weather variables such as pressure, temperature, humidity and velocity. There are no analytical solutions for these equations; rather they must be approximately simulated on supercomputers. The related integrals and derivatives are calculated directly on the grid describing the area of calculation or with serial approximations as in spectral or finite element models. [73]

The basic equations describing the behavior of atmosphere are the Newton's second law, the law of conservation of mass, the state equation of an ideal liquid, the first law of thermodynamics and the law of conservation of humidity. They can be written

as follows [73]:

$$\begin{aligned}
\frac{\mathrm{d}\vec{V}}{\mathrm{d}t} &= -\frac{1}{\rho}\nabla p - 2\vec{\Omega}\times\vec{V} - \vec{g} + \vec{F} \\
\frac{\partial\rho}{\partial t} &= -\nabla\cdot(\rho\vec{V}) \\
p &= \rho R T \\
Q &= c_p\frac{\mathrm{d}T}{\mathrm{d}t} - \frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}t} \\
\frac{\partial\rho\kappa}{\partial t} &= -\nabla\cdot(\rho\kappa\vec{V}) + \rho S
\end{aligned} \tag{33}$$

where we denote by $\vec{\Omega}$ the angular velocity, $\vec{F}$ the turbulent and viscose friction, Q the non-adiabatic net source of heat, $\vec{V}$ the wind velocity, $\vec{g}$ the gravitation, T the temperature, R the gas constant, S the water steam net source and by c_p the specific heat of dry air at constant pressure.

The NWP models must be initialized using a known set of states describing the atmosphere. This state information is retrieved using a various number of measurements, coming from weather stations, radiosondes, satellites etc. Then in order to interpolate the data to cover all areas of the Earth the meteorologists use a procedure called *data assimilation*. This data is used for instance in the *European Centre for Medium-range Weather Forecasts* (ECMWF) model which models the weather in all over the globe. [64]

This information provided by the ECMWF and local observations is fed into a *local area model*. Examples of such models are the *High Resolution Limited Area Model* (HIRLAM) and *Application de la Recherche à l'Opérationnel à Méso Echelle* (AROME), which have a 7.5 km and 2.5 km grid length, respectively. More specifically, it is the local observations that give the information of the state of the atmosphere inside the HIRLAM/AROME grid and the ECMWF that provides the boundary conditions to those models for simulation. [36]

Both the world-wide ECMWF and locally centralized HIRLAM/AROME models are, as stated before, a set of equations (as in (33)), which are discretized and solved both in place and time. The discretization length gives the limit for the accuracy for such models: it is not possible to model phenomena that are active on smaller scales than the grid resolution [64]. However, as the grid resolution after the application of local area models is still rather coarse for purposes of very locally oriented weather characterization and forecasting needed in wind power applications, the meteorologists refine the weather predictions by using a procedure called *parametrization*. One of such models is called *Wind Atlas Analysis and Application Program* (WAsP). In the case of the Finnish Meteorological Institute, it gives a final grid resolution of 250 meters [36]. The parametrization consists of describing local thermal effects, turbulences, landscape roughnesses etc. [64]. It is therefore essentially a statistical model based on empirical facts written in the language of mathematics about the

area where the very local weather is predicted.

The forecasting length of the meteorological models described above is between 48–172 hours ahead. They are usually run every 3–6 hours [64]. According to [86], a higher operational resolution could improve the accuracy of forecasts slightly, but incur an expensive cost. This fact thus imposes the same temporal resolution to the wind power forecasting methods as well, while normally the wind power forecasting tools employ the NWP data as their inputs.

4.2 Wind power forecasting

Wind power can be forecast in various different time scales. The choice of an appropriate time scale depends on the needs for which forecasting is performed. Typical needs are summarized in Table 1 [41].

Table 1: Wind power forecasting time scales

Time scale	Forecasting needs
Seconds	Turbine control
Minutes/Hours	Regulation, corrective markets
Day(s)	Day-ahead market commitment
Months	Resource allocation
Years	Economic viability

As one can see in Table 1, the first time scale of seconds corresponds mainly to wind turbine control, which can be enhanced by using wind forecasts on a very short time-scale instead of historical winds measured after the turbine rotor. For further references on these issues, we ask to refer to e.g. [6]. Respectively, the two longest time scales of typically more than a month correspond to long-term variations of the wind, which have mainly planning, developmental and investment related implications. In this thesis, the scale at which wind and wind power are predicted corresponds to minutes and hours. In these time scales, it has been shown that wind speed forecasts play an essential role in the performance of power predictions [29].

How then to transform the obtained wind forecast into power forecast? The easiest answer to this question is the use of the wind turbine power curve (Figure 1). It is supplied by the wind turbine manufacturer and gives the mapping between wind speed and generated power; however, there are several inconveniences in using such a curve. The first one is quite straightforward: if a wind farm consists of several wind turbines manufactured by different turbine manufacturers, how then to combine the different curves into one describing the total production of the wind farm? The second problem is however much more severe: as the power curve is obtained using measurements in wind tunnels, it can only be seen as a theoretical reference.

As seen in Section 2.1, the wind does not have the same energy before entering and after leaving the wind turbine. Assume that we have a large wind farm with

many turbines approximately in a row. If the wind blows from the end of such a row, the last turbine of the row will experience a much lower wind speed than the first turbine did. This is called the *wake effect* and it directly influences the power curve of the ensemble of wind turbines. There are methods for estimating such wake effects (see e.g. [19]). In any case, the presence of the wind wake effect means also that the wind direction plays a role in wind turbine power generation. There may be also some other effects, such as the air temperature, humidity and even a wind speed cross-influence [15], which can have an impact on the air density (see (5)). Thus it cannot simply be envisaged, that the power curve of a wind farm could be solely dependent on the wind speed prevailing in the wind farm area.

However, there is an essential factor in determining the produced power of a wind farm, which so far has been neglected. This factor is the availability of wind turbines in the park. The total power output of a wind farm depends naturally on the number of wind turbines being *online*, i.e. participating actively in transforming wind into energy. Consequently, in the forecasting procedure one has to somehow model and forecast the number of available turbines. This issue will not be addressed in this thesis: it is rather just taken into account by purifying the effect of available turbines in the data set. This means that the simplest approach is to say that final power forecast will be a product of the number (or total capacity) of turbines available in the future and the power forecast of a *virtual wind turbine* representing the characteristics of the wind farm.

Given the facts presented above, we can now turn our attention towards formulating a generic wind power forecasting problem. It is essentially a mission to successfully describe the wind farm future production based on the information of the current and past states of a set of some variables. In this thesis, all the variables are of discrete nature, which means that they are sampled in time.

Let there be a set of observed wind farm production values $\{P_t, t \in \mathbb{T}\}$, where $\mathbb{T}$ is the ensemble of time indices and P_t the production value at time instant t. Knowing the evolution of P up to time instant t, we want to predict the future value of P at time $t+k$. Here k is called the *forecasting horizon*. Put in words, we might say that

> given the *information set* and assuming that the identified behavior of the process continues in the future, we can say that... [64]

The challenge of forecasting is just here: we implicitly make the assumption that the so far observed behavior of the process $\{P_t\}_{t\in\mathbb{T}}$ will continue to be the same in the future. This crucial assumption poses great challenges in keeping the prediction models up-to-date.

Now the general forecasting procedure can mathematically be put in the following form:

$$\begin{aligned} P_{t+k|t} &= g(P_t, ..., P_{t-l_P}, \xi_t, ..., \xi_{t-l_\xi}, \hat{\chi}_{t+1|\tau}, ..., \hat{\chi}_{t+l_\chi|\tau}, \hat{q}_{t+k|t}) + \epsilon_t \\ &= g(\Phi_t', \hat{q}_{t+k|t}) + \epsilon_t \end{aligned} \tag{34}$$

where ϵ_t represents random (white) noise and the information set can be written as

$$\Phi_t^{'} = \left[P_t, ..., P_{t-l_P}, \xi_t, ..., \xi_{t-l_\xi}, \hat{\chi}_{t+1|\tau}, ..., \hat{\chi}_{t+l_\chi|\tau}\right] \tag{35}$$

The problem is thus to find the best approximation for function g, which maps the information set containing the previous total production values of the wind farm, the previous and predicted values of sets of external variables $(\boldsymbol{\xi}, \hat{\boldsymbol{\chi}})$ and the forecast for the number of available turbines $(\hat{q})$ to the forecast of total power generated at a future instant $t + k$. One can note, that such model formulation belongs to the class of *Nonlinear AutoRegressive eXogenous (NARX)* models.

Let us now introduce a little transformation in order to facilitate the use of real wind farm data (see Sections 5.3 and 6). As discussed above, the total wind farm power generated at any instant, denoted here by P_t, requires the knowledge of the number of available turbines at the same moment. Put in another way, we cannot estimate the total produced power without knowing the number of turbines involved in the process. If we let

$$p_t = \frac{P_t}{q_t \cdot P_{nom}} \tag{36}$$

we have $p_t \in [0, 1]$ at every instant t. The power p_t expresses the power generated by a *virtual turbine* representing the whole wind farm. In this way the modeling can be guided at directly searching the best function f satisfying

$$\begin{aligned} p_{t+k|t} &= f(p_t, ..., p_{t-l_p}, \xi_t, ..., \xi_{t-l_\xi}, \hat{\chi}_{t+1|\tau}, ..., \hat{\chi}_{t+l_\chi|\tau}) + \epsilon_t \\ &= f(\Phi_t) + \epsilon_t \end{aligned} \tag{37}$$

where

$$\Phi_t = \left[p_t, ..., p_{t-l_p}, \xi_t, ..., \xi_{t-l_\xi}, \hat{\chi}_{t+1|\tau}, ..., \hat{\chi}_{t+l_\chi|\tau}\right] \tag{38}$$

The function f could be called the *wind farm power forecasting curve*, which accentuates on its role in forecasting the future power of the wind farm based on a set of observed and predicted information.

The performance of wind power forecaster can be assessed in many ways. Normally one defines a measure ϑ which is used in assessing the performance of forecasting machinery. Examples of such measures are the *mean absolute error* (MAE) and the *root mean squared error* (RMSE), which are defined as

$$\begin{aligned} \vartheta_{\text{MAE}} &= \frac{1}{N} \sum |\hat{p}_t - p_t| \\ \vartheta_{\text{RMSE}} &= \sqrt{\frac{1}{N} \sum (\hat{p}_t - p_t)^2} \end{aligned} \tag{39}$$

As normally a WPP will test numerous different forecasting approaches, it usually wants also to rank them according to their accuracy. In such a process usually a simple forecaster is used as a reference. One may introduce an improvement metric, which for any (continuous) measure is be defined as

$$\iota_\vartheta = \frac{\vartheta^{(ref)} - \vartheta}{\vartheta^{(ref)}} \tag{40}$$

The measures of accuracy (and improvement) may of course be functions of e.g. forecasting lead-time. In any case, the standard measures of errors (39) take into account both the systematic and random part of forecast errors. If one wishes to further study the individual impact of those to the total error, the RMSE may be decomposed [43]. Let us first recall that $e_t = \hat{p}_t - p_t$, so for $\vartheta_{\mathrm{RMSE}}$ one has

$$\begin{aligned} \vartheta^2_{\mathrm{RMSE}} &= \frac{1}{n}\sum_t e_t^2 = \frac{1}{n}\sum_t e_t^2 - (\frac{1}{n}\sum_t e_t)^2 + (\frac{1}{n}\sum_t e_t)^2 \\ &= E[e_t^2] - E[e_t]^2 + E[e_t]^2 = Var[e_t] + E[e_t]^2 = Var[e_t] + \mathrm{bias}^2 \end{aligned} \tag{41}$$

It is possible to further decompose the variance term in (41) so that finally one obtains [34]

$$\vartheta^2_{\mathrm{RMSE}} = \mathrm{bias}^2 + \mathrm{sdbias}^2 + \mathrm{disp}^2 \tag{42}$$

where with the notations $\sigma(x)$ as the standard deviation of x and $r_{x,y}$ as the correlation coefficient of x and y, the terms of (42) are

$$\begin{aligned} \mathrm{bias} &= \frac{1}{N}\sum(\hat{p} - p) \\ \mathrm{sdbias} &= \sigma(\hat{p}) - \sigma(p) \\ \mathrm{disp} &= \sqrt{2\sigma(\hat{p})\sigma(p)(1 - r_{\hat{p},p})} \end{aligned} \tag{43}$$

Furthermore in [43] it has been shown that a *phase error*, which is quite common in NWP predictions, manifests itself (mainly) in the dispersion (disp) component of (42). This is because the dispersion measures the cross-correlation of predicted and measured power. On the other hand, the bias and sdbias components together indicate amplitude errors in prediction. They can theoretically be corrected from the forecast by linear correction (substracting the bias and scaling by the inverse of the standard deviation bias), but the dispersional bias cannot, as correlation is a linearly invariant measure.

Efforts for such corrections exist in the literature, e.g. in [49], where a method for reducing systematic errors has been presented: Kalman filtering techniques have been applied in order to estimate the state of bias according to historical (recent) measurements and then correcting the original NWP by this estimation of bias.

As normally the accuracy of a model gets better when plugging in more explanatory variables, some metrics measuring the model generalization ability have been developed. The most famous one is perhaps the *Akaike information criterion* [1], which measures the goodness of a model fit. The goodness can be understood to be a sort of trade-off between the complexity and accuracy of the model. The criterion, which is to be minimized, for a model with k parameters based on a data set with n observations and fitting residuals $\hat{e}$ is written as

$$AIC = 2k + n(\ln(\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2)) \tag{44}$$

4.3 Literature review of wind power forecasting methods

This Subsection presents some wind power forecasting methods and implemented models based on recent research. They are published in various journals, including e.g. *Applied Energy, Renewable Energy, IEEE Transactions on Power Systems, Electric Power Systems Research* and *Wind Energy*. Neither the list of publications nor the following list of models is by no means exhaustive. We rather try to present several different modeling approaches and their corresponding abilities to capture and mimic something essential related to wind power, and further to forecast its behavior in time. Other literature reviews on different modeling approaches can be found in [18, 86, 44, 46]. The implemented forecasting models in Section 6 will be based on the found literature results.

There are two fundamental problems in wind power forecasting. The first one is the transformation of given numerical weather data into power output, and the second is the subtle selection of relevant other historical factors impacting the power in future. To tackle these problems, we can identify (at least) five different types of modeling approaches. The first one can be called a *naïve reference method*, as it is exaggeratedly simple and does not consider weather variables at all. The next two are the *physical method* and the *statistical method*, which will be described below. Furthermore, there are also *ensemble/hybrid* and *probabilistic* forecasting methods.

4.3.1 Naïve reference methods

Wind is fundamentally a process, which changes its mean and variance all the time [86]. Statistically meaningful patterns of some kind, which normally are easily identified for instance in electricity consumption as people in general behave in the same way given a day type, are not as easy to distinguish with wind speed data. On the other hand wind has a rather long memory and thus its values correlate with previous ones with a significant order. It could be imagined that a very short-term behavior of wind power could be modeled with the following property:

$$\hat{p}_{t+k|t} = p_t \tag{45}$$

Such approach is called *persistence forecasting* and it simply assumes that the behavior of wind continues in the future exactly as it was up to the time when it was observed the last time. As could be expected, its performance degrades as the lead-time k gets bigger; however, it can be surprisingly difficult to beat in very short-time horizons (from minutes to 1 hour). Thus the persistence model serves as a good reference tool for comparing performances of other more complex model structures. Furthermore, it is to be noted that if the lead-time $k = 1$ and if the prediction is considered to be the expectation of future given the past, (45) equals the famous martingale [20] property. It simply argues that the best prediction of wind or wind power one step ahead is its present value.

4.3.2 Physical methods

The NWP models provide predictions on a rather large grid resolution, corresponding mostly to the needs of meteorologists. The NWP models are not run every hour and their grid resolution can be quite unsatisfactory for the needs of a wind farm; consequently, all the local effects may not be taken into account. Hence the output of those models needs to be further refined to the very specific level of the wind farm in question. This extrapolation procedure is usually called *downscaling*. For carrying out this, a detailed specification of the terrain of the wind farm must be available. The downscaling models take extensive use of sophisticated flow modeling techniques based on *Computational Fluid Dynamics* (CFD) [87]. However, collecting the information regarding the terrain characteristics of the wind farm may be very time-consuming and is also subject to errors, which directly affects the modeling result [86].

When all the local effects, including obstacles, roughness, orography and wake effects are taken into account, the physical models eventually produce an estimate of exact prevailing winds at the wind farm co-ordinate and hub height. Then there are two additional tasks to perform. The first one is to plug the wind estimate into a power curve transforming it into power output. The second one is usually called *Model Output Statistics (MOS)*, which tries to remove systematic forecast errors by applying regression techniques on the estimate and the observed output. The use of such models naturally necessitates the use of a *Supervisory Control And Data Acquisition* (SCADA) system providing online measurements of realized production (or wind speed).

There are several operational physical models used by wind power producers and TSOs. One of them is *Previento*, which is largely used for example in Germany [27]. It provides estimates on future production of a single wind farm, but also for a larger area where many wind farms are situated. For this purpose, it uses an *up-scaling* algorithm [44] in order to take account the spatial smoothing effect discussed in Section 2.2.

4.3.3 Statistical and AI methods

If the physical methods aim at describing the direct relationship between the fuel and the production, i.e. the wind and power output of a wind farm at a specific location, the approach can be somewhat different as well. If one tries to directly model the mapping between past production values, enhanced with a set of other explanatory variables and the future power output, those models belong to a collection called *statistical* or *Artificial Intelligence* (AI) methods. These types of models employ relationships between variables that may have no meaningful physical correspondence. This is especially true for the AI model class. Even though the physical phenomena are not really taken into account, expertise of the problem is naturally strongly appreciated in finding the best variables explaining the behavior of the process.

All statistical models are based on fitting a predefined structure to a data set, known usually as the *training set*. The fitting is thus an optimization procedure, which aims at finding the best mapping function (37) minimizing the predefined cost or *loss function*. The loss function is usually the sum of squared errors between the model and the real observed outputs. It can be shown that such minimization translates to the forecast of the process being an estimate of the conditional expectation of the output, given the information set and the chosen model [16]. As the parameters are unknown but the model is predefined, the classical statistical models are known as *grey-box models*. In the AI modeling paradigm such constraints on structures can to some extent be relaxed, and they are usually characterized as being *black-box models*. Then as the structure and its parameters are estimated, one normally validates the model by using a *validation set*, which is unseen to the model contrary to the training set used in estimation. In order for the model to be performant, it has to capture the behavior of the validation set as well. Later on the measures of accuracy are all calculated over the validation set.

The overview presented in this thesis is by no means exhaustive, as there are surprisingly many types of statistical models in the literature, all of which succeed at various rates in explaining the future of wind speed or production. We will simply adapt a strategy going from the very simplest approaches to somewhat more intricate ones.

The conventional and simplest statistical models are directly based on linear time series analysis. They require a set of historical observations (on either wind speed or production) and then try to establish a mathematical model describing the relationship of past and future. Undoubtedly, the classic in this field of science is the Box-Jenkins approach for modeling linear time-series [12]. It describes an *Autoregressive Integrated Moving Average* model ARIMA(p, d, q), which for wind power output p_t can be written as follows:

$$\varphi(B)\Delta^d p_t = \theta(B)\epsilon_t \tag{46}$$

where ϵ_t is *white noise*, $\Delta^d = (1-B)^d$ the difference operator and the lag polynomials

of order p and q are defined as

$$\begin{aligned} \varphi(B) &= 1 - \sum_{i=1}^{p} \varphi_i B^i \\ \theta(B) &= 1 + \sum_{i=1}^{q} \theta_i B^i \end{aligned} \tag{47}$$

Such models forecasting wind and related power can be found e.g. in [79]. As can be seen from those references, the performance of such models is quite mediocre, and it deteriorates very quickly after a few hours, slightly depending on the wind farm site characteristics. The more stable the atmosphere is in the location of a wind farm, the easier it is to capture a meaningful statistical relationship and model it further. In any case, the reasons for such timeseries models as in (46) to fail are evident: they fundamentally require a process, that is both linear and has constant moments (mean and variance). Thus they cannot satisfactorily describe the behavior of wind speed (see Figures 12, 13).

The next step to consider is to take external variables into account, and the most obvious choice in wind power forecasting is naturally the wind speed's NWP. Actually according to [49], today it is hardly envisaged that performant models would not use any NWP data, because using only autoregressive information it is not possible to follow any sudden changes in wind speed. They rather just follow a local trend, as in pure autoregression-based forecasting no new information becomes available after starting the forecasting engine.

There is also another choice to make: normally wind power is forecast multiple steps ahead. There could be thus two possibilities: either to use a single model and iteratively run it up to the preferred forecasting horizon, or to design a separate model for each horizon. The alternatives each have their benefits and drawbacks. In a single model approach the errors are cumulated as the iteration goes on, since after the initial forecasting step only predictions on the power output are available. On the other hand, it can be rather unobvious to use production information of, let's say time step $t-6$, to predict production at instant t. In Section 6 both approaches are considered.

As the computing power has increased steadily over the past years, new modeling possibilities requiring a lot a computational power have emerged. Some examples of such models, which have been used in wind speed and power prediction, are *Artificial Neural Networks* (ANN), *Fuzzy logic / Adaptive Neuro-Fuzzy Inference System* (ANFIS) models and *Support Vector Machines* (SVM) and their combinations. Such models typically require a large amount of historical data, which is used in tuning their internal structure and parameters. There is however another side of the coin: they are all subject to overfitting problems arising from using too complex a structure with regard to information contained in data.

An illustration of overfitting is presented in Figure 8, where a *Multilayer Perceptron* or *FeedForward-BackPropagation* (FF-BP) network is used to estimate a functional

relationship from noisy and erroneous measurements. The real relationship is captured by the equation $\sin(t)(1-t^2)\exp^{-t^2}$. The measurements are obtained by corrupting the function output by a Gaussian noise with a standard deviation of $\frac{1}{50}$. Some observations are also altered to get an impression of measurement freezing. One can see, that although the errors between the observed and simulated outputs are constantly decreasing as one increases the number of ANN neurons (increasing model complexity), the relationship between X and Y clearly becomes less interpretable. Noisy and even erroneous observations become modelled with the real data as well, which is rarely the desired result. Because of this very typical behavior care must be taken in the pursuit of finding the best forecaster f in (37) when using AI methods.

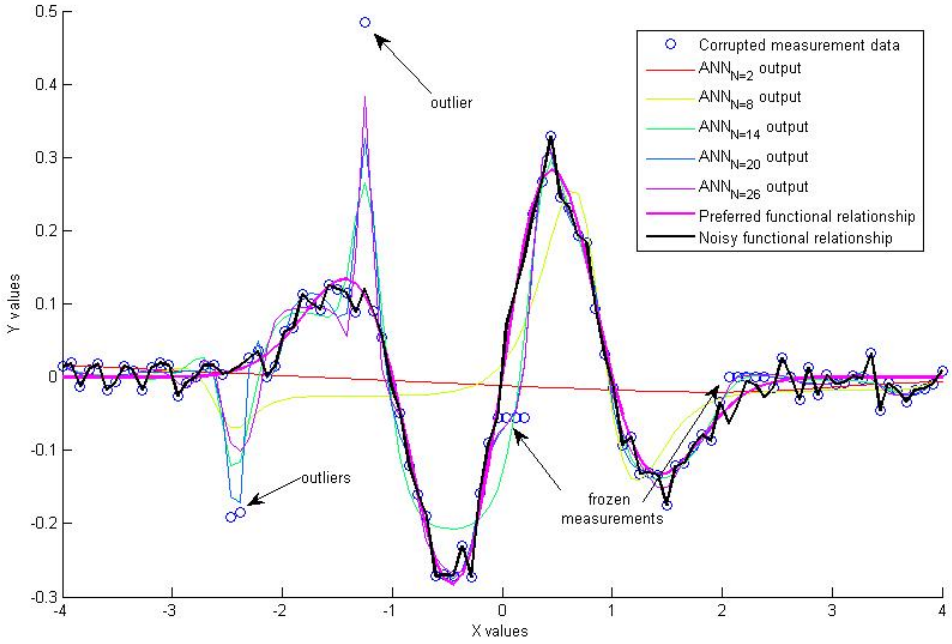

Figure 8: Neural networks trained with noisy data and different number of neurons N.

When using data coming from measuring instruments, care must also be taken in filtering out possible measurement errors. It may prove to be useful to divide the observed time series (of wind speed or production) into components, which describe the behavior of the observed process at different scales. In a way, such procedure could be imagined to be a sort of mathematical magnifying lens through which one obtains a clearer picture of the fundamental characteristics of the process in question. With the same kind of approach one is also able to more effectively remove measurement noise and errors.

Wavelet analysis is a relatively new mathematical technique, which has been used in various signal processing purposes in describing a signal at various scales of interest. Lately, in [45, 24, 42] they have been used in wind speed prediction as well. We here

briefly introduce wavelet analysis following those sources.

The main reasons for using wavelet-based approaches in signal processing are that wind speed is not stationary, so that within the realm of traditional time series approaches one does not attain a sufficient level of accuracy. Using wavelets it is possible to decompose the observed time series into approximately stationary components and then model them separately. No initial assumptions on linearity or stationarity of the time series have to be made. Finally the aggregate forecast may be obtained as a simple summation of the separately predicted components.

A wavelet transform resembles the traditional well-known Fourier transform in the way that it reveals frequency-based facts of a time-based signal. It is also based on applying a sort of base function to the signal. In Fourier analysis these bases are sines and cosines. Nevertheless, the Fourier transform has a significant limitation, as it loses completely the time information related to the measured signal when transformed into frequency domain. The wavelet transform overcomes this limitation by using a variable-sized windowing technique. It allows to simultaneously investigate long and short time intervals, where low and high frequency information is observed. The basis functions are not limited to predefined ones as in Fourier analysis – the wavelet approach rather uses an infinite set of possible basis functions. The purpose of basis functions are however similar: they provide a way to break the original signal into components of several frequencies by projecting them into translated and scaled versions of the basis, called more often *mother wavelets* $\psi(t)$. Examples of mother wavelets are the *Haar* and *Daubechies* wavelet families.

A continous wavelet-transform of original signal $f(t)$ according to scale a and translation b[1] is defined as

$$W_f(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t)\bar{\psi}(\frac{t-b}{a})dt \tag{48}$$

where $\bar{\psi}$ is the complex conjugate of the mother wavelet.

It has however been shown that the continuous transformation is redundant, as not all the coefficients $W_f(a,b)$ are needed in recovering the original function. Actually, by choosing *dyadic* scales and translations ($k, j \in \mathbb{Z}$) as

$$\begin{aligned} a &= 2^{-j} \\ b &= k \cdot 2^{-j} \end{aligned} \tag{49}$$

the discrete wavelet transform is just as accurate as the continuous counterpart.

In order to effectively study the original signal at different levels for identifying specific patterns at a specific scale, one applies a *multiresolution analysis* technique. It expresses the original signal as a sum of a constant vector A and J other vectors D_j. By using the Mallat algorithm [52], the discrete multiresolution-based wavelet transform is implemented using low- and high-pass filters. The approximation can

[1] Apart from this part, these symbols are reserved for other purposes

be characterized to be a low-frequency representation of a signal, whereas the detail is a substraction of two approximations at successive levels. The general trend of the original signal is held in the approximation and the high-frequency components in the details.

Wavelet decomposition based on Mallat's approximation and details is presented graphically in Figure 9. Illustration on the wavelet decomposition of wind speed measured over a 400 hour horizon is presented in Figure 10. The decomposition is done using a Daubechies3 wavelet at level 3.

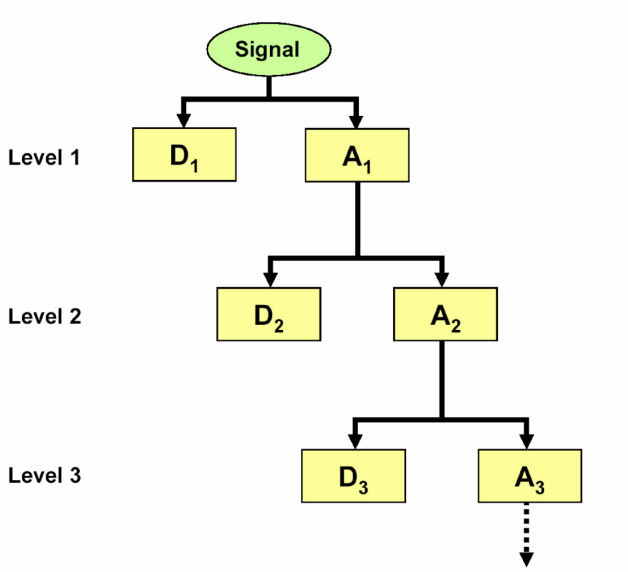

Figure 9: Wavelet decomposition tree into approximation (A) and details (D). Source: www.biomedcentral.com/

4.3.4 Hybrid and ensemble methods

The object of the so-called *hybrid model* is to benefit from the advantages of each of the two main modeling approaches described earlier. Using both physical and statistical methods, it can be possible to achieve both interpretability and flexibility, respectively. Usually such integrated models take advantage from the higher accuracy of statistical models in shorter horizons and the reliability of physical NWP methods, which take into account local phenomena. In this thesis such models are not further considered per se, while the physical approach requires a significant amount of expert knowledge of e.g. CFD methods. In literature, descriptions of hybrid models can be found e.g. in [18].

More generally, one can naturally embed several different models of any type and weigh them somehow in order to produce an accurate forecast. Such models are usually called *ensemble models*, which insists on the role of using many alternative

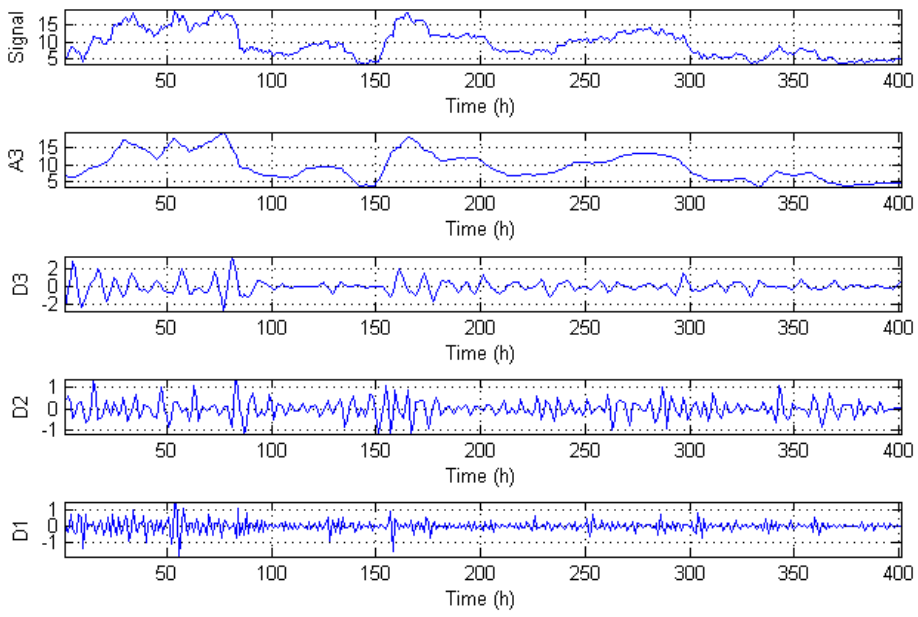

Figure 10: Measured wind speed (on y-axis, m/s) decomposed into approximation and details

predictions. As big WPPs usually procure several NWPs in order to reduce the inherent uncertainty concerning wind speed, such models have gained more and more attention in last years. Ideally, the resulting ensemble forecast should be better or at least as good as the best individual forecast.

How is the combination of predictions carried out? While the ensemble predictions form today an essential part of the state-of-the-art wind power prediction models, the idea of combining forecasts dates back from the 1969. The first contribution to this research was done by Bates and Granger, who published an article describing several methods for combining a set of forecasts [9]. In their publication, the past errors of individual forecasts are used to determine the optimal weight in averaging the forecast into an ensemble one. Several different methods for deriving these weights are presented. More recently, the combination methods are elaborated for the wind power forecasting case in [72, 58].

The following is closely derived from the discussion in [72]. If a WPP has K competing predictors, which comprise the ensemble set

$$\boldsymbol{p}_{t|\tau}^{ens} = \left(p_{t|\tau}^{(1)}, p_{t|\tau}^{(2)}, \ldots, p_{t|\tau}^{(k)}, \ldots, p_{t|\tau}^{(K)}\right), t < \tau \tag{50}$$

the related past errors of a predictor k observed up to instant τ can be written as

$$e_{\tau|\tau-l_p}^{(k)} = p_\tau - \hat{p}_{\tau|\tau-l_p}^{(k)} \tag{51}$$

Now the combination procedure is aimed at building a (linear) combination of the individual predictions:

$$\hat{p}^C_{t|\tau} = \sum_{k=1}^{K} \mu^{(k)}_{t|\tau} \hat{p}^{(k)}_{t|\tau} \tag{52}$$

The best combination $\hat{p}^C_{t|\tau}$ minimizes a predefined loss function $L(\cdot)$ for $e^{(C)}_{t|\tau} = p_t - p^{(C)}_{t|\tau}$, so that we finally have the weights

$$\mu^{(k)}_{t|\tau} = \arg\min L(e^{(C)}_{t|\tau}) \tag{53}$$

If the loss function is defined as $L(e^{(C)}_{t|\tau}) = E\left[(e^{(C)}_{t|\tau})^2\right]$, the realized power output as $p_t = \boldsymbol{\mu}^T_{t|\tau}\hat{\boldsymbol{p}}_{t|\tau} + e^{(C)}_{t|\tau}$ and furthermore the weights are constrained such that $\sum_k \mu^{(k)} = 1$, the solution of the combination problem can be expressed as

$$\boldsymbol{\mu}^*_{t|\tau} = \frac{E^{-1}\left[\boldsymbol{e}_{t|\tau}\boldsymbol{e}^T_{t|\tau}\right]\mathbf{1}}{\mathbf{1}^{\boldsymbol{T}} E^{-1}\left[\boldsymbol{e}_{t|\tau}\boldsymbol{e}^T_{t|\tau}\right]\mathbf{1}} \tag{54}$$

Note that the weights determined by the condition (54) minimize the variance of combination forecast assuming it is unbiased [58]. In reality, the covariance matrix of errors $E\left[\boldsymbol{e}_{t|\tau}\boldsymbol{e}^T_{t|\tau}\right]$ has to be replaced by its estimate $\hat{\boldsymbol{V}}_{t|\tau}$, as no future errors corresponding to instants after τ are available at the moment of combination. A possibility for this estimation is to use an adaptive *Exponentially Weighted Moving Average* (EWMA) procedure, which gives weight to recent observations according to equations

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{t+1} &= \gamma\hat{\boldsymbol{\mu}}_t + (1-\gamma)\hat{\boldsymbol{e}}_{t+1} \\ \hat{\boldsymbol{V}}_{t+1} &= \gamma\boldsymbol{V}_t + (1-\gamma)(\hat{\boldsymbol{e}}_{t+1} - \hat{\boldsymbol{\mu}}_{t+1})(\hat{\boldsymbol{e}}_{t+1} - \hat{\boldsymbol{\mu}}_{t+1})^T \end{aligned} \tag{55}$$

In such a way the covariance of individual predictor errors gathered in $\hat{\boldsymbol{e}}_t$ can be estimated online. For instance, suppose one has two competing predictors, whose corresponding error variances (up to some time instant) can be collected to a matrix

$$\hat{\boldsymbol{V}} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \tag{56}$$

The optimal combination weights according to (54) are written as

$$\boldsymbol{\mu}^* = \begin{pmatrix} \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \\ \frac{\sigma_1^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}} \end{pmatrix} \tag{57}$$

The weights are hence directly determined by the variances of the competing predictors' errors.

A possible drawback of the presented weight calculation method is that it implicitly assumes that the estimates for error mean and covariance up to forecast instant τ are valid later on at the actual forecasting time steps $(\tau + 1, \ldots, t)$ as well. It has also been criticized that such combination of individual predictors all based on a same information set (a single NWP information for instance) cannot generally outperform the best individual predictor. In such a case a better alternative could be to pick the best individual from the set of competing ones, for which techniques with the above criticism have been proposed in [72]. On the other hand, if a WPP relies to many alternative NWP providers the weighted averaging technique presented above generally yields better forecasting results [58].

4.3.5 Probabilistic methods

All of the methods presented above aim at providing accurate forecasts on future power generation. Their internal structures and inputs vary, but they all put out a single number corresponding to power (or energy). As explained in Section 3.4, such *point forecasts* do not tell anything about their variability. This can be illustrated with a simple example: if we have a set $A = \{2, 3, 4, 5, 6\}$ and another set $B = \{3, 3.5, 4, 4.5, 5\}$ they both have a common mean 4, but B has a lower variance. The same reasoning applies to wind power forecasting if the sets A and B now are imagined to be data to which forecasters have been fitted: neglecting the uncertainty every forecasting model inherently has may have serious implications when the results of such a model are applied to trading electricity.

According to [39], there are two different methodologies related to wind power uncertainty characterization, which both have different subclasses as well. The first one tries to augment point predictions provided by an existing wind power forecasting system with uncertainty information. This type of probabilistic forecasting is called *prediction error approach* in the literature. The second approach aims at directly providing WPP with probabilistic predictions on future power. We will briefly discuss examples of both types of systems found in literature.

The simplest approach to providing uncertainty information is to compute for instance the standard deviation of forecast errors related to different forecasting look-ahead times [39]. Due to the chaotic nature of climate, however, in order to be accurate it is needed to take into account different circumstances under which the forecasting errors occur. This can be done by separating the errors into classes according to some variables, of which a natural candidate would be the predicted power. Using such methodology WPP can obtain situation-dependent uncertainty assessments of their forecasting procedure. Many authors in the literature consider *quantile regression* as a suitable method for such a purpose.

If one wants to completely (re)design a WPP forecasting system to account for the needs of probabilistic forecasting, a method for constructing the probability density

function for wind power has to be applied. In order to be fully flexible, it cannot be limited to any predefined quantiles or intervals (i.e. pairs of quantiles). It must be chosen whether to use a predefined distribution form (like Gaussian for instance) or whether to use more generic non-parametric techniques like *kernel density* estimation theory. An example of the latter can be found at [39]. In either case, the goal is to find a conditional probability density function mapping the information available at time τ to a later forecasting instant t.

The use of quantile regression (QR) provides an attractive way to map predicted amount of production into realized one. As discussed in Section 3.4, in a simple optimization paradigm the strategic day-ahead bid may be defined as a certain quantile of future power distribution. Even if not aiming at making optimized bids according to (29), knowledge of production forecast uncertainty may prove fruitful.

We briefly discuss the properties of quantile regression following [57]. In quantile regression a certain quantile $Q(\omega), 0 < \omega < 1$ is expressed as a linear combination of some known regressors and unknown coefficients, which are estimated from the data. Having r different regressors $\boldsymbol{x}$, this may be written as

$$Q(\omega, \boldsymbol{x}) = \beta_0(\omega) + \beta_1(\omega)x_1 + \cdots + \beta_r(\omega)x_r = \beta_0(\omega) + \sum_{j=1}^{r} \beta_j(\omega)x_j \tag{58}$$

Now the coefficients are estimated based on observations $(y_i, x_{i,1}, \ldots, x_{i,r}), i = 1, \ldots, N$. In ordinary linear regression, one has a loss or *check function* $\rho_\omega(e) = e^2$. Replacing it with

$$\rho_\omega(e) = \begin{cases} \omega e, & e \geq 0 \\ (\omega - 1)e, & e < 0 \end{cases} \tag{59}$$

one obtains the vector coefficient estimates

$$\hat{\boldsymbol{\beta}}(\omega) = \arg\min_{\beta} \sum_{i=1}^{N} \rho_\omega \left(y_i - (\beta_0(\omega) + \beta_1(\omega)x_{i,1} + \cdots + \beta_r(\omega)x_{i,r}) \right) \tag{60}$$

If one wants to allow a more sophisticated approach by using some basis functions f (for instance splines [11], which are used in [48]) instead of linear combination as in (58), we may write

$$Q(\omega, \boldsymbol{x}) = \beta_0(\omega) + \sum_{j=1}^{r} f_j(x_j; \omega) \tag{61}$$

With the quantiles estimated, it must be checked whether the distribution they predict is reliable: in other words, whether the QR gives results that are approximately in line with the observed quantiles calculated from the data. Furthermore it should be checked whether these estimations are robust regarding the absolute value of

explanatory variables as well as their location in the training data [48]. If this is the case with an acceptable level of accuracy, they can be used in the exact way they are defined: at first one plugs in the relevant input variables (predicted power at t issued at $\tau < t$, say) and the nominal probability ω. The output of quantile function then tells the level *below* which the realized production will fall with the nominal probability.

Such probabilistic methods described above aim at providing an a priori warning of prediction uncertainty. However, in an operational context such warning in the form of probabilities supplied by e.g. quantile forecasts may be somewhat esoteric. In [68] another possibility for providing forecast related risk information has been developed. The technique called *skill forecasting* is actually a blend of ensemble and probabilistic methods, while it aims at informing the operators of wind power forecast about its possible uncertainty using ensemble predictions.

A *Normalized Prediction Risk Index* (NPRI) from forecasting look-ahead time k_1 to k_2 is defined as

$$\begin{aligned} \mathrm{NPRI}(k_1, k_2) &= \frac{1}{k_2 - k_1 + 1} \sum_{i=k_1}^{k_2} \tilde{\sigma}_{t,k} \\ \tilde{\sigma}_{t,k} &= \left[\frac{J}{J-1} \sum_{j=1}^{J} \mu_j (\hat{p}^{(j)}_{t+k|t} - \hat{p}^{(C)}_{t+k|t})^2 \right]^{\frac{1}{2}} \end{aligned} \tag{62}$$

which effectively measures the dispersion of wind power predictor ensembles (see (50), (52)) over a set of successive look-ahead times. It can be shown [68] that the relation between NPRI and prediction errors can be expressed in a probabilistic manner using contingency tables, where prediction errors are divided by their range and put together with their corresponding NPRI values. The tables (in a form of Box plot for instance) thus reflect the conditional probability of power imbalances given NPRI values.

5 Case study on a real wind farm

This Section provides a case study on a wind farm situated in Kemi, Northern Finland. It is owned and operated by PVO-Innopower, a subsidiary of Pohjolan Voima, and consists of 10 wind turbines of 3 MW power each. Thus the total power of the park is 30 MW, making it today the largest wind farm in Finland. From now on, the Finnish wind power producer shall be referred as FWPP, if necessary.

This Section is divided as follows: the collection procedure of the measurement data is first described. Then using the collected data we analyze and discuss the characteristics of wind speed measured at the site. Subsequently, the currently used wind power forecasting procedure is presented and its performance discussed.

5.1 Data collection

At the time of writing this thesis, there is no centralized SCADA system in the representative wind farm. However, all ten wind turbines in the farm possess a measurement unit and a database. The databases can be accessed in order to retrieve the generated power as well as the measured wind and its direction. What can be seen as the first serious deficit is that there is no measurement mast for the realized wind speed and directions. While the wind turbines measure the wind with a cup anemometer [6], which is located after the rotor plane, the measurement is automatically biased: as the wind gives some of its energy to the rotating turbine rotor, its speed slows down and is not the same as before entering the turbine (see Section 2.1). Nevertheless as there are no such individual measurements we had to use the data coming straight from the turbines.

The data in the database is measured every 2 seconds, and then averaged over a 10 minute interval. Also a maximum and minimum value for the 10 minute intervals are recorded, but they are not used in our models. Because the operating unit of electricity markets as well as wind forecasts is one hour, the 10 minute values are further averaged to an hourly value. Then the final problem is to somehow average the 10 individual turbine measurements to a single value representing the whole farm. Before discussing the procedure for this averaging, the reader may note that such averaging is of course by no means absolutely necessary: one could build several models for a single farm, describing for instance distinct clusters of the farm. As the size of the farm is currently quite small and for the sake of simplicity, only one value corresponding to the whole farm was deemed relevant and sufficient.

The averaging over the individual turbines is made as follows: firstly, the number of turbines available at each 10 minute interval is determined. There is no binary variable available corresponding to a turbine being in operation or not. For this

reason, we use a simple heuristic for determining the operational status of a turbine:

$$q_t = \begin{cases} 1, & P_t^{min} > 0 \wedge P_t > 0 \\ 0, & \text{otherwise} \end{cases} \tag{63}$$

The definition (63) for turbine availability at moment t (now still at 10 min scale) states that a turbine has been online, if its minimum production as well as its production over a 10 minute period have been positive. This condition takes into account the fact that with very low winds, there may be short periods (2 seconds would be enough) over which a turbine produces something, but the whole 10 minute average is zero (as it is actually a net average reflecting the slight consumption of power as well). Furthermore, (63) cannot distinguish whether a turbine has been broken or not producing because of too low winds. Such information is not necessary for our applications, while the final power output of the whole farm is defined by (36), which accounts for the final output and wind measurement only turbines which are available. Put differently, the final measurements are obtained using only turbines which are operational. In this way, the average values can be calculated using 1–10 turbines, depending on q_t. It is possible that for some instant $q_t = 0$, which means that either there is not enough wind for turning the turbines or all the turbines are out-of-use. Such time steps are removed from the data set in order to not have excessively zero values harming the further model estimation.

For wind speed, the averaging is done similarly as for the power output except that it is not normalized; however for the wind direction the calculation is somewhat more intricate as direction is not continuous as a measure. This is why a different approach has to be considered by calculating the individual sines and cosines of each measured angle, then averaging them over the available turbines and finally by calculating the average angle. This procedure can be written as follows:

$$d_t^{(total)} = \arctan\left(\frac{\frac{1}{N}\sum_{i=1}^{N} q_t \sin d_t^{(i)}}{\frac{1}{N}\sum_{i=1}^{N} q_t \cos d_t^{(i)}} \right) \tag{64}$$

Finally, it must be noted that it may be necessary to further scale the result of (64) in order for it to span the whole region $[0, 2\pi)$, but this depends slightly on the computer implementation of arctan.

5.2 Wind characteristics at a real site

In Section 2.2 we have presented the theoretical Weibull distribution used to derive statistical assumptions on wind speed behavior. Figure 11 presents the Weibull fit made to 1-hour averaged data collected over an approximately one year horizon. One can clearly see that the wind speed distribution does not resemble a typical Weibull distribution very well as there is a strong peak of medium-range (4–8 m/s) winds in the data series. As we identified several distribution candidates, a *Birnbaum-*

Saunders distribution seems to model the statistical behavior of measured wind speed reasonably well. However the data set is slightly biased as it must be kept in mind that wind measurements, with which every turbine has been offline, are deleted from the data. For information, the Birnbaum-Saunders distribution is characterized by two coefficients β, γ[2] and a corresponding PDF

$$f(w) = \frac{1}{2\pi} \exp\left\{ -\frac{(\sqrt{\frac{w}{\beta}} - \sqrt{\frac{\beta}{w}})^2}{2\gamma^2} \right\} \left(\frac{(\sqrt{\frac{w}{\beta}} + \sqrt{\frac{\beta}{w}})}{2\gamma w} \right) \tag{65}$$

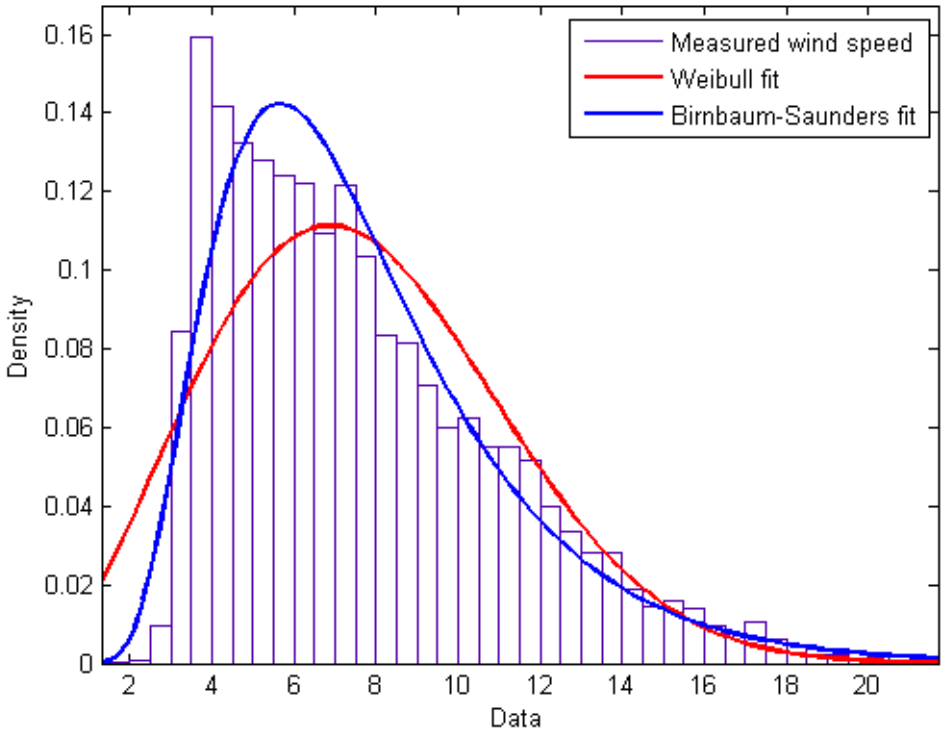

Figure 11: Measured wind speed and statistical fits

It has been claimed in various publications (e.g. [41, 23]) that the wind speed time-series is a realization of a so-called *long-memory process*. Such processes have a rather intuitive feature suggested by its name: the coupling between values of the process at successive instants is stronger than in ordinary processes, which may be called having a *short-range dependence*. There are several ways to put that kind of a phenomenon into mathematical form, but one of the most intuitive methods is to use autocorrelations. It can be argued that contrary to the case of short-memory processes, whose autocorrelation function typically either drops to zero after a certain lag or at least approaches zero exponentially, the autocorrelation functions

[2]Apart from this part, these parameter symbols are reserved for other purposes

of long-memory processes typically exhibit a power-like decaying behavior. These phenomena are more generally related to concepts like *self-similarity* and even *chaos*, which will not be further discussed here [50].

However, the interesting thing about such long-memory processes is that such characteristic of wind speed would allow us to use specific methods (*Autoregressive Fractionally Integrated Moving Average* (f-ARIMA or ARFIMA)) models [41]) developed for modeling those kind of processes. This is why in Figures 12 and 13 we present the autocorrelation and partial autocorrelation functions [12] of the measured wind speed, starting at two different random time steps. This allows us to investigate the possibility of long-range dependence. Generally, such analysis is valuable per se, as it may reveal something essential of the process, which helps in modeling and prediction its behavior.

In Figure 12 the autocorrelation function of measured wind speed, starting at two randomly chosen instants of time, is presented. Additionally, Figure 13 shows the autocorrelation function for the first difference of measured wind speed, i.e. $\Delta^1 w_t = (1 - B)w_t = w_t - w_{t-1}$ and the partial autocorrelation function for the measured wind.

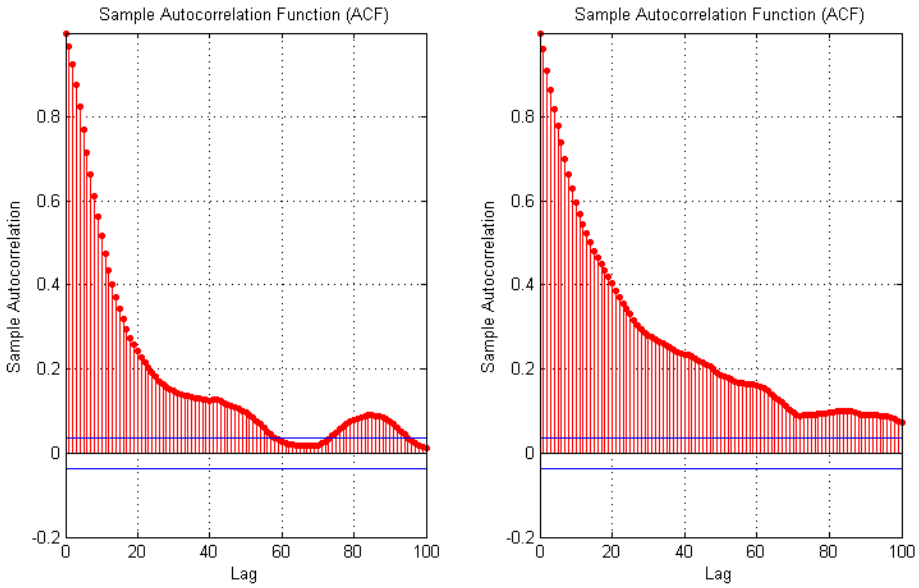

Figure 12: Autocorrelation functions of measured wind speed

Analyzing the Figures 12 and 13 it can be seen that the wind speed really has a rather long memory. However, (partial) autocorrelation function actually is defined only for stationary processes, which the wind speed clearly is not (see Figure 15). It can nevertheless be used in process characterization bearing in mind its interpretability limitations.

One can try to obtain stationarity (of mean) by differentiating the time series of wind speed once. Investigating its autocorrelation function we note that it is zero

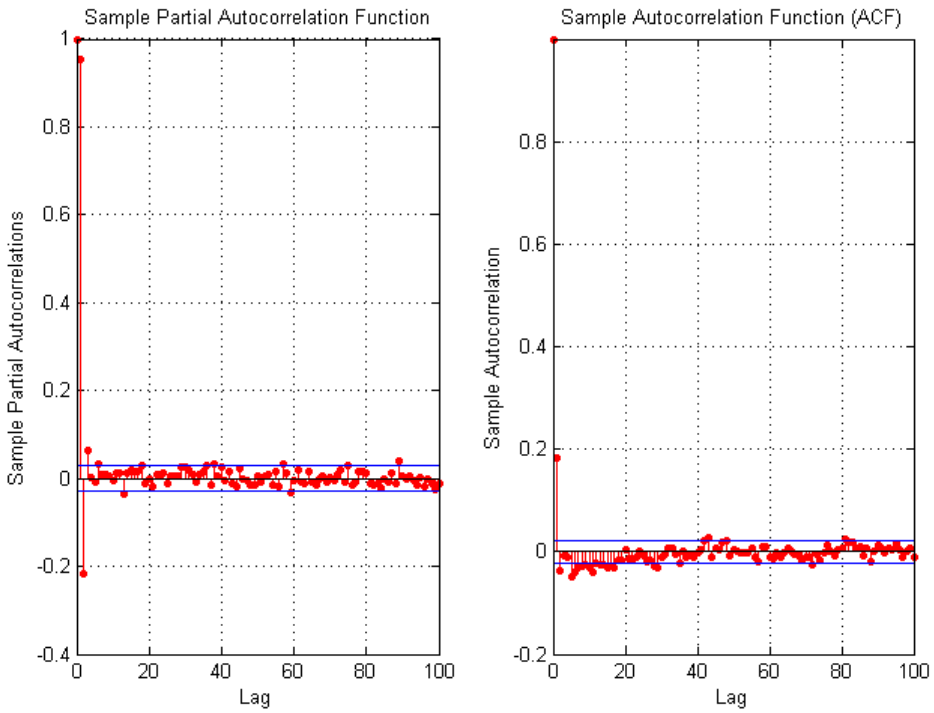

Figure 13: Partial autocorrelation function of differenced wind speed (left) and autocorrelation function of differenced wind speed (right)

at every lag above one with a 95% confidence level. This suggests that a MA(1) model could be used. The partial autocorrelation function of the differenced wind speed calculated at several regions of the data suggests that indeed all lags after 2 (or perhaps 3) are insignificant, so that AR(2) or AR(3) models could be attempted. Together this implies that perhaps fitting an ARIMA(2/3,1,1) (see (46)) model could be attempted, although recursive techniques should be preferred in order to better account for non-stationarity.

Turning the attention from the magnitude of wind to its direction, some conclusions on wind behavior can also be made. It must be acknowledged that even if theoretically the wind direction plays no role in turning the kinetic energy of wind into power, due to reasons described in Section 4.2 it may have some value, at least in large wind farms.

The left part of Figure 14 presents a *wind rose* of measured wind direction. It shows an angular histogram of wind speed. It seems that wind mainly blows from every direction, but there is a slight concentration of Western wind and especially of the ones coming from directions of order 200-240 degrees. It is in line with the location of the wind farm facing the Bothnian Bay on its Western side. It seems also that the winds blowing from the sea are stronger than the ones coming from the continent. This can be seen from the right part of Figure 14, which presents the measured wind speed as a vector quantity according to its magnitude and direction.

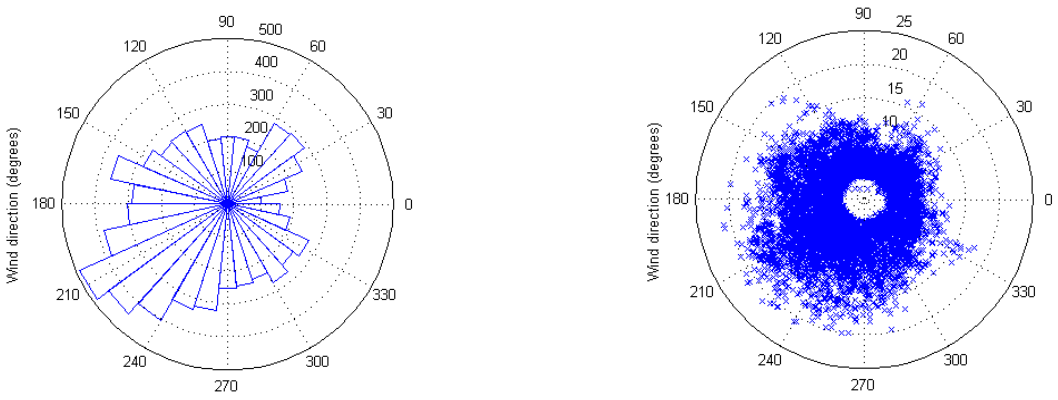

Figure 14: Wind direction histogram and the distribution of wind speed by its direction

The stationarity (in a weak-sense) of recorded wind speed time-series can be investigated by calculating the means and variances of the time-series at various time windows. For this purpose we divided the speed record into 12 successive windows, consisting of 743 hours (approximately one month) each. Then we plot the mean and variance of each window. The results together with their bootstrapped [21] 95% studentized confidence level limits are presented in Figures 15 and 16. They clearly reveal the non-stationarity of wind speed at the wind farm site, while both the mean and variance alter between the successive time windows with a statistical significance.

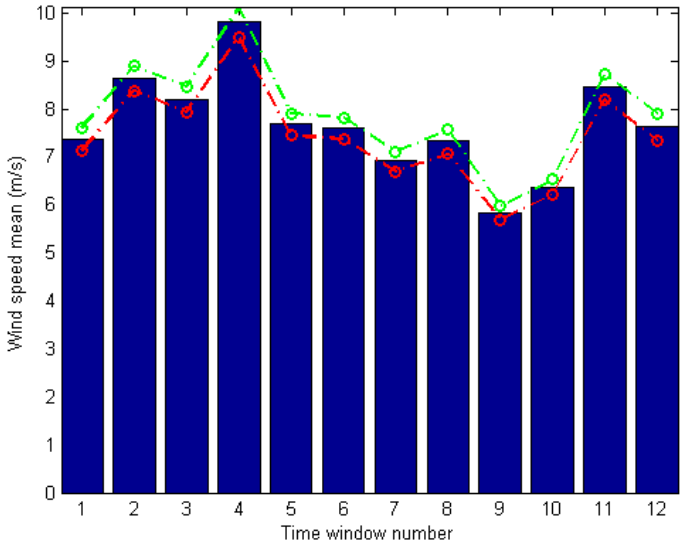

Figure 15: Wind speed means with their 95% confidence limits at different time windows

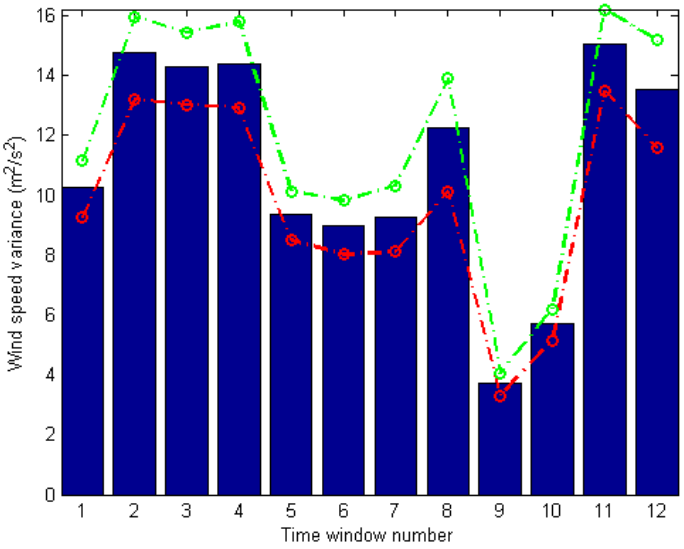

Figure 16: Wind speed variances with their 95% confidence limits at different time windows

5.3 Evaluation of current wind power forecasting procedure

Today, the wind power prediction of the FWPP is solely based on a NWP, which it gets once a day from a certain meteorological institute. It is issued usually at around 8 a.m., which can be said to be its basetime. However its values start at 10 a.m. and they cover the present day and the next day. Thus the forecast horizon of the NWP is 3–40 hours, and the lifetime of a single NWP vector is 38 hours. The bids for the next day are based on wind forecasts with a horizon of 17–40 hours ahead, while the bids are submitted around noon. Lately, the FWPP also uses the newest NWP to correct its power commitments for the day in question, that is using the NWP covering the forecasting range 3–16 hours ahead. Figure 17 clarifies the power forecasting procedure form the NWP point of view.

In order to transform the obtained NWP into power, a simple power curve representing the whole wind farm is used. It is written as

$$\hat{P}_{t|\tau}(\hat{w}_{t|\tau}, \hat{q}_{t|\tau}) = \begin{cases} \hat{q}_{t|\tau} \cdot \min\left\{1, \frac{1}{3} f(\hat{w}_{t|\tau})\right\}, & 4 \leq \hat{w}_{t|\tau} \leq 21 \\ 0, & \hat{w}_{t|\tau} < 4 \vee \hat{w}_{t|\tau} > 21 \end{cases} \tag{66}$$

where

$$\begin{aligned} f(\hat{w}_{t|\tau}) = &-0.0015 * (\hat{w}_{t|\tau} - 2)^3 + 0.0564 * (\hat{w}_{t|\tau} - 2)^2 \\ &-0.0725 * (\hat{w}_{t|\tau} - 2) + 0.1207 \end{aligned} \tag{67}$$

Figure 18 presents forecasting curve (67) graphically. From (66) and (67) it is clear

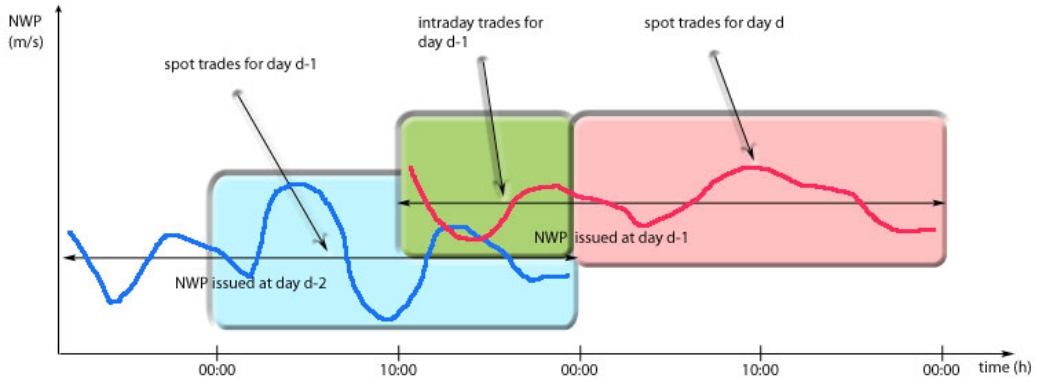

Figure 17: Currently used NWP timeline

that the outputs of the current prediction procedure are total power forecasts using the wind speed prediction for every hour. The results are in megawatts (MW), and with the assumption that constant power is produced during one hour, they can be directly transformed into energy (MWh).

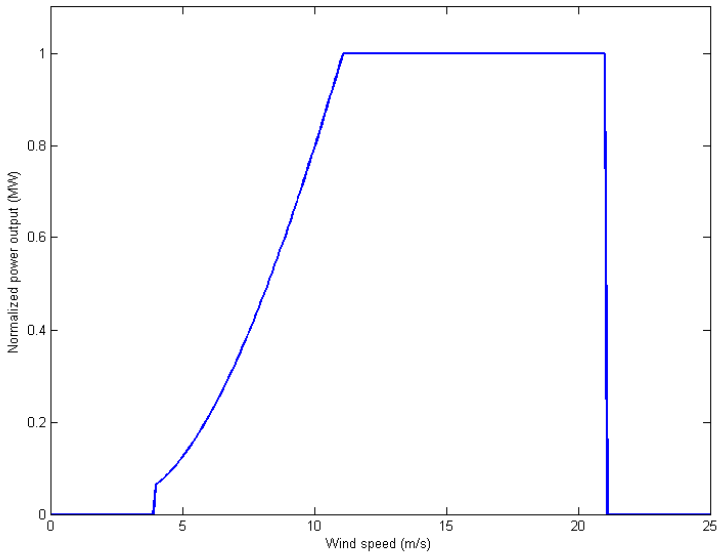

Figure 18: Reference forecasting power curve according to (66)

Now in (66) one can see that the accuracy of the power prediction can be divided into two subcategories: the one related to the predicted number of available turbines and the one related to the NWP. Although the accuracy of correctly predicting the number of available turbines – or put in another words, the endeavor to keep them running all the time, is clearly essential, there is not much operators can do about it. It seems reasonable that the best (and simplest) prediction regarding the

turbine availability is the current condition of the turbines plus their expected future condition (announced future maintenance works). In the presence of strong wind, the danger of turbine shutdown becomes imminent and the operators might consider lowering the availability index as a precaution. We will not discuss predicting the availability any further in this thesis.

The NWP accuracy can be seen as the most important factor in assessing the overall accuracy of power predictions. The measured wind and production as well as the predicted wind is depicted in Figure 19, which is a snapshot of the wind farm over 336 hours (2 weeks) between late May and early June 2009. It is evident that the measured wind corresponds very strongly with the measured production. However, the correlation of predicted and measured wind speed is only approximately 76%. This can be read from Figure 19 as well, where the biases of NWPs are clearly visible and can be quite large. The same relationship can be seen from Figure 21 where the NWP values are plotted against the measured wind speed. Clearly there is significant variation in the forecast accuracy. Note that the granularity of NWPs is only 1 m/s.

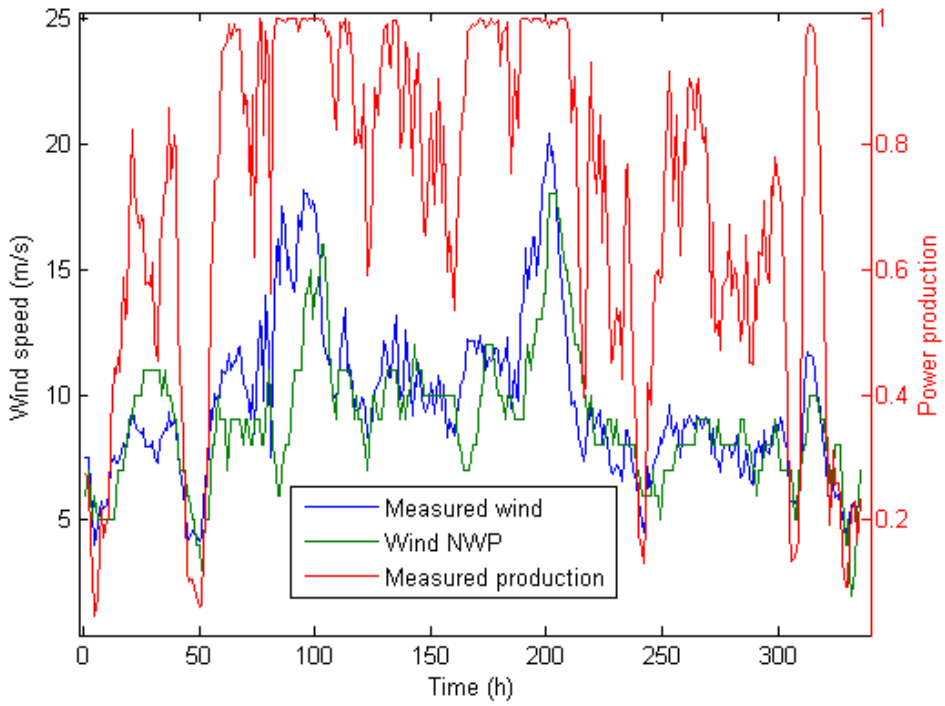

Figure 19: Measured wind, its prediction and measured power output between late May and early June 2009

The predicted power and the real measured power can be illustrated graphically. For this purpose we omit the turbine availability index appearing in (66) and (67), as we are only interested in the normalized output. Thus the possible inaccuracies are directly the NWP's inaccuracies, but they may be fortified or dampened depending

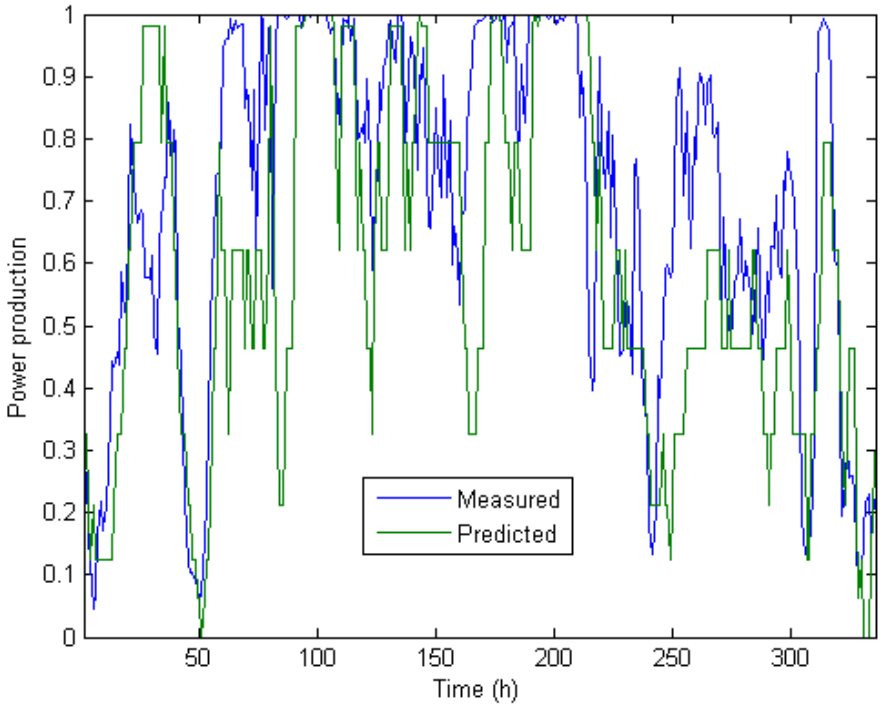

Figure 20: Power forecast using a power curve and measured power between late May and early June 2009

on the magnitude of the measured wind. Figure 20 presents the predicted and measured power during the same time interval as in Figure 19.

As can be clearly seen, the errors related to inaccurate information of the wind at the wrong time (see also Figure 3) can lead to significant errors and thus to economical losses. Furthermore as the initial day-ahead prediction is not actively corrected based on the *observed* information on production and wind speed, the errors pertain and cumulate. There is thus a need for a method to refine the predictions as approaching the time of physical delivery. This will be further developed in the next Section.

Lastly we may want to analyze the effect of NWP error on the power output error. As the effect of turbine availability has been removed from the data (see (36)), the two remaining sources of error in wind power prediction with the current forecasting model are the NWP error and the error in transforming NWP into power prediction. Looking at Figure 19 it is quite evident the NWP error dominates here. The further question is: does the NWP bias change when the wind speed prediction becomes older?

From Figure 17 we can see that currently the NWP used in power prediction is released only once a day. Intuitively one would suggest that an older NWP, i.e. one corresponding to instants that are longer away from the releasing base-time, could be more inaccurate than a more recent one. They are based on a set of equations,

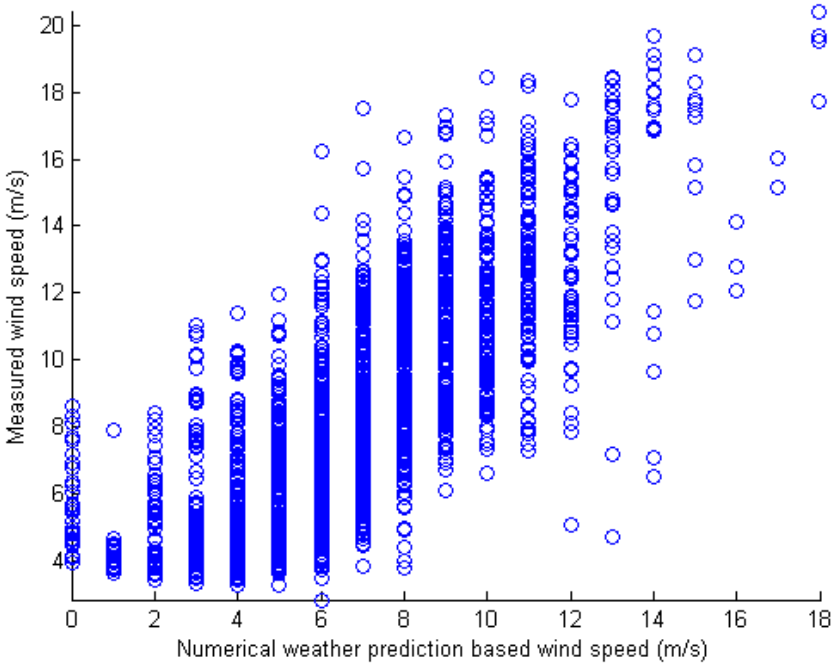

Figure 21: NWP vs. measured wind speed

which need a set of initial states in order to calculate the trajectories of relevant weather variables ahead in time. Thus the error could accumulate as time goes on.

At first, we have divided the wind speed error, defined as

$$e_t = \hat{w}_t - w_t \tag{68}$$

according to the corresponding hour $h \in [0, 23]$. By *bias* we refer to the error expectation. The left part of Figure 22 presents the mean absolute errors $|e_t|$ calculated at all different hours in the whole dataset. In the right part the corresponding sample error variance $Var[e_t]$ is shown. In addition, the 95% studentized confidence limits of these estimates have been calculated using a bootstrapping method in order to point out possible outliers in the data set.

The interpretation of Figure 22 is rather ambiguous: it seems that at low NWP look-ahead times (after 10:00) the absolute error is at its minimum, as is the error variance. However, with longer lead-times the situation does not significantly change after the first rise at around 13:00. Curiously, it even seems that the NWP absolute error decreases slightly when its lifetime is almost at end. The same behavior can be read from Figure 23, which depicts the evolution of NWP error distribution. The histogram has been augmented with a fitted Gaussian distribution. It seems that the hourly errors may be satisfactorily considered as Gaussian.

The pure means of NWP errors (and correspondingly the means of fitted Gaussians in Figure 23) are presented in Figure 24. It shows that no clear pattern of NWP bias

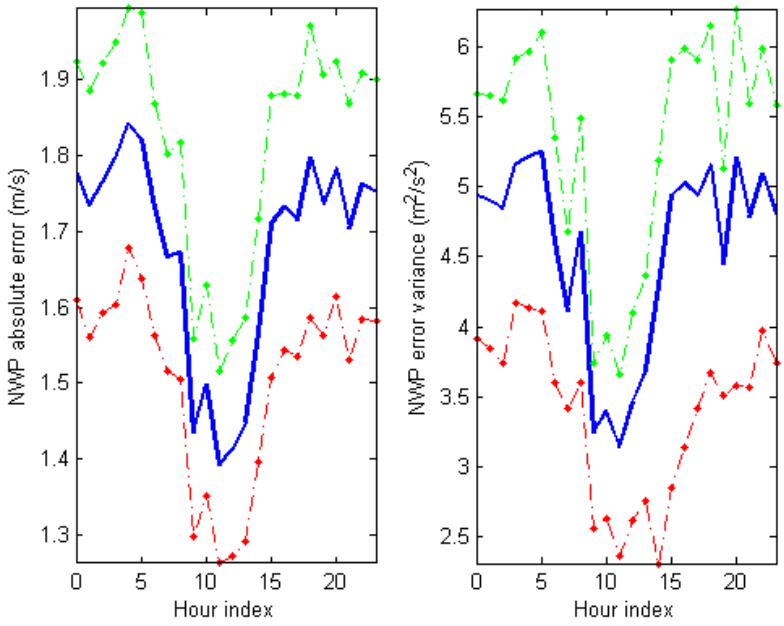

Figure 22: NWP mean absolute error and error variance with their 95% confidence limits by hour

can be determined. It rather seems that the hourly errors are not directly linked to the age of NWPs.

However, if the NWP bias is plotted against its own value we obtain quite a different result. It clearly seems that the current wind speed prediction underestimates the true wind, when the predicted value is low or medium. In high predicted values the situation is opposite: the true wind is statistically smaller than the predicted one and thus the NWP bias turns from negative to positive. There are not, however, many points corresponding to large NWPs, so conclusions concerning these values must be drawn carefully. This can be seen as the confidence limits getting wider. The results are presented in Figure 25.

We also estimated the RMSE of the wind speed prediction and decomposed it into terms corresponding to (42). The results shown in Table 2 show that the biggest inaccuracy is due to phase errors. It means that the provider of NWPs has been inapt to calibrate its weather model exactly to the co-ordinate of the Kemi wind farm. There are also significant amplitude errors in the prediction.

Table 2: NWP RMS error (m/s) decomposed

RMSE	2.2404
bias	-0.6840
sdbias	-0.1082
disp	2.1308

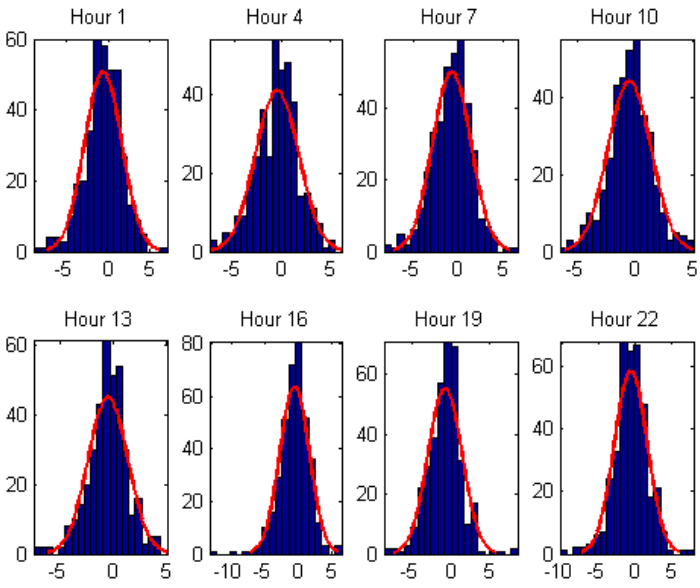

Figure 23: Histogram of NWP errors conditioned on the prediction hour

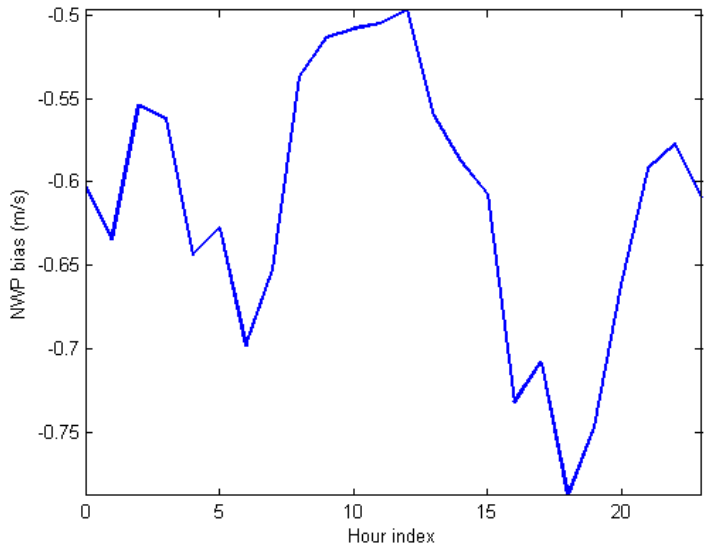

Figure 24: NWP bias by hour

Finally, looking at Figure 26 we see that even perfect wind forecasts do not lead to perfect power forecasts, as given a specific wind speed value, there is significant hourly variation in generated power. The main reason for this is probably the hourly variation of wind speed itself: one must keep in mind that the measured values are hourly averages, so that wind speed may fluctuate inside the hour and due to power curve nonlinearities the fluctuations are either dampened or amplified. The values in Figure 26 are hourly averages.

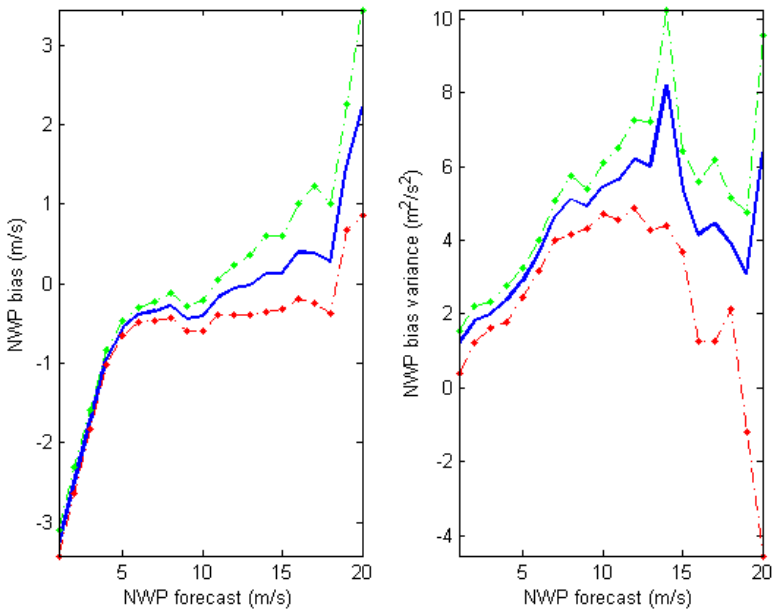

Figure 25: NWP bias with their 95% confidence limits against the NWP value

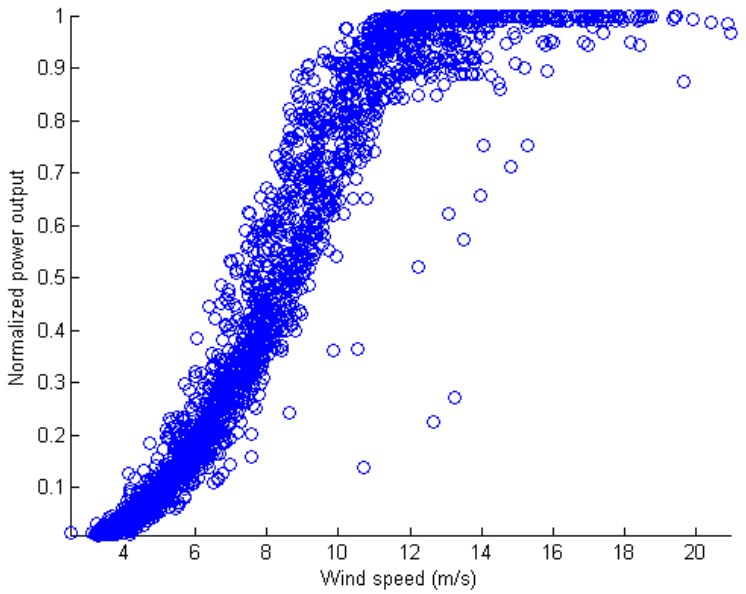

Figure 26: Empirical power curve of the Kemi wind farm

6 Implementation of an advanced wind power forecasting procedure

In this Section, new advanced methods for wind power forecasting are presented. The proposed methods are two-fold: at first, the initial day-ahead power scheduling procedure is to be enhanced by introducing a new model for transforming wind forecasts into power forecasts. Secondly, the models have also the emphasis of correcting the initial day-ahead power commitments by re-forecasting the power output for time horizons of 1–6 hours. This correction then helps the WPP to cope with imbalances with regard to initial power forecasts and realized output.

The Section is divided into three parts. In the first one, we discuss general guidelines in the forecaster design. The second one presents a mathematical description of the implemented models. Thirdly, some implementation issues are addressed.

6.1 Design paradigms

As seen in Section 4, there is a vast number of wind power forecasting models described in the literature. What differentiates them from each other is their internal structure and the inputs they employ. There are nevertheless also some common characteristics shared by all investigated models: their accuracy is far from perfect and it deteriorates as the forecasting look-ahead time increases. These facts are of course very intuitive, but in assessing the need for designing an advanced forecaster must be kept in mind. No perfect forecaster has (until now) been designed, and the most important factor related to wind power forecasting remains to be the quality of NWPs used almost always as a forecasting input. As normally the expertise of a WPP does not lie on meteorology, it has to rely on the provider(s) of NWPs and then try to employ them as efficiently as possible.

In our pursuit towards designing an advanced wind power prediction model, some general guidelines were identified: firstly, the model had to be as accurate as possible given the available data. Secondly, it had to be implementable, meaning that a principle of parsimony was respected: given two equally accurate models, the simpler one is preferred. While the wind power producer aims at using the implemented model operationally in the future, the model also had to be easily maintainable in order to account for new fresh data, and the possibility to use it in other wind farms. Thirdly, it was acknowledged that a single model is not suitable for handling all situations: as already seen in the literature review, using forecaster ensembles one normally can reduce the forecasting error. The forecasting accuracy is a function of the look-ahead time also in a sense that an accurate model with short look-ahead times may behave worse than another model when dealing with longer forecasting horizons.

Based on the findings in the literature it became quite evident that an AI-type model is the closest one matching the above requirements. Such models share many

common features, including that they are relatively straightforward to implement and maintain and that there is a lack of need for physical modeling, which requires expert knowledge on fluid dynamics. On the other hand, AI models need a rather large set of historical data, which in addition must not be heavily corrupted as it deteriorates the accuracy of the model very quickly. The lack of physical relationships may also be somewhat confusing, so that care must be taken in interpreting the forecasting results. Generally such models are far from being a forecasting crystal ball, but nevertheless they have proven to be suitable and efficient in wind power forecasting purposes. As will be seen, this proves to be true in our case as well.

In [47] it has been shown that generally there is no clear winner when competing different AI methologies against each other in wind speed or power forecasting. This is verified by an empirical study in our case as well. While normally one cannot directly choose a universally superior state-of-the-art model and base the whole forecasting procedure on implementing it alone, a more flexible approach had to be chosen. It was concluded that, based on literature findings, a couple of models are implemented and tested and then based on those testing results the final model (ensemble) is chosen.

At first, the available data has to be divided into two distinct sets, where the first one represents the history that is presented to the candidate models. They learn the essential relationships from it, and then their generalization capability is tested and validated using the other data set. The validation is done using a RMSE metric (39), with the aim of minimizing the validation RMS error and simultaneously respecting parsimony, i.e. not choosing an overly complex model.

The available historical data ranging from 5.9.2008 11:00 to 30.10.2009 10:00 (after pruning 8136 hours) consists of variables presented in Table 3. Note that the time instant t has been dropped for simplicity of notation.

Table 3: Available data used in model estimation and performance testing

Variable	Forecasting	Economics
total production P	x	
normalized production p	x	
available turbines q	x	
wind speed w	x	
wind direction d	x	
wind speed forecast $\hat{w}$	x	
wind direction forecast $\hat{d}$	x	
spot price λ_d		x
elbas price λ_i		x
balancing prices (up, down) $\lambda_{-/+}$		x

Originally, the measured variables P, p, q, w and d are on a 10 minute sampling basis and the rest of the variables are gathered hourly. Because nowadays the smallest time unit used in the Finnish (or Nordic) electricity markets and with NWPs is

essentially an hour, all the variables should be resampled to an hourly basis. To account for possible measurement noise and outliers, a wavelet-filtering approach using a Daubechies4 wavelet at level 4 has been at first tried in reducing the noise in measured production and wind, respectively. However, it was concluded that a simple hourly averaging would dampen the possible measurement noise and variation almost as effectively. Furthermore, we did not succeed in finding a suitable wavelet candidate for production values, whose bounded nature posed difficulties on the boundaries of the filtered values. Based on this, the wavelet-filtering approach was abandoned and a simple hourly averaging was employed for both wind speed and production. An illustration on the filtration effect in the wind speed time series is presented in Figure 27.

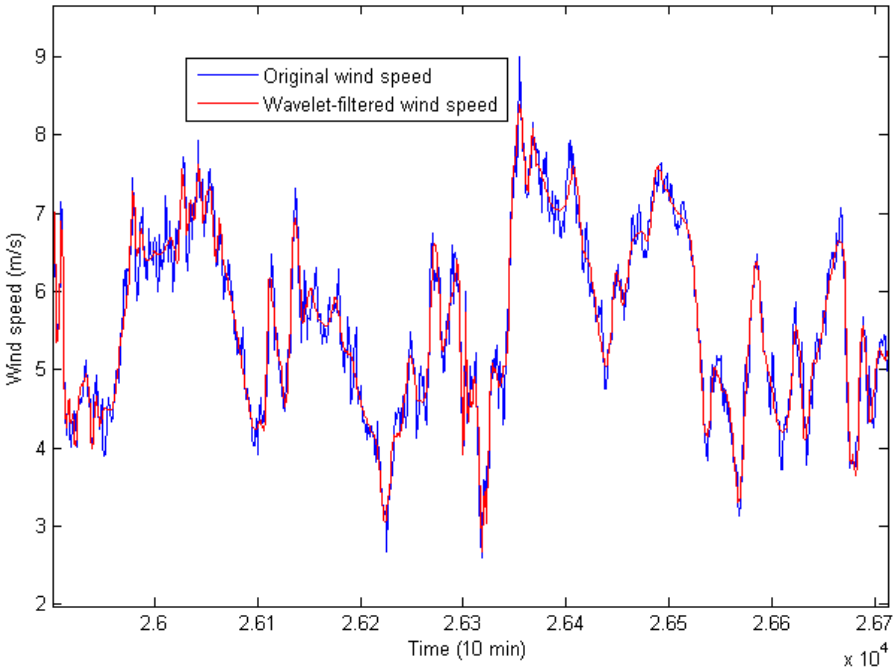

Figure 27: Wavelet-filtered wind speed (10 minute sampling)

6.2 Description of the models

Based on some preliminary usability tests on various modeling approaches, the set of model candidates was limited to the following. The abbreviations in the following list correspond to *Feed-Forward BackPropagation (Multilayer Perceptron)* and *Radial Basis Function* neural networks, *Least-Squares Support Vector Machine*, *Adaptive Neuro-Fuzzy Inference System* and *AutoRegressive (with eXogenous inputs)* models.

In Table 4, the four first candidate models are typical findings in the literature describing AI-based wind power forecasting. The wavelet-AR(X) model is used in estimating an own short-term forecast for wind speed, in order to account for

Table 4: Candidate forecasting models

Model	Inputs	Output	Forecasting horizon (hours)
FF-BP	$p, \hat{w}, \hat{d}$	$\hat{p}$	1-36
RBF	$p, \hat{w}, \hat{d}$	$\hat{p}$	1-36
LS-SVM	$p, \hat{w}, \hat{d}$	$\hat{p}$	1-36
ANFIS	$p, \hat{w}, \hat{d}$	$\hat{p}$	1-36
Reference power curve	$\hat{w}$	$\hat{p}$	1-36
Wavelet-decomposed AR(X)	$w, \hat{w}$	$\tilde{\hat{w}}$	1-6

possible systematic biases in original NWPs. The last model used as a reference is defined by (66).

One must note that although in all the advanced models in Table 4 the inputs are defined to consist of an equal set of variables, the models may differ in the way they are used. More specifically, the construction of all individual models is a search for the best model in two dimensions: firstly, in the internal structure, which e.g. in the case of feed-forward networks means the number of neurons and the type of activation functions. Secondly, the input space must be specified, meaning e.g. how many lags of historical production values are taken into account. Altogether this means that the search space one theoretically should explore is vast. In order to fully be able to search for all (meaningful) candidates, one should use some kind of heuristic approaches, as for instance cross-validation based on genetic programming. Such approach is used in the literature in e.g. building models for electricity load forecasting [4], and for wind speed prediction in [26]. However, as the budgeted time did not allow us to build such an approach for testing a large number of models, the candidates are selected by going through a rather limited subset of possible models by hand. This kind of approach could be called *trial and error* and it may well be performant enough in finding (sub-)optimal forecasting models.

For the description of FF-BP, RBF and ANFIS models regarding to their internal structure, training and validation methods and their general characteristics we ask the reader to refer to e.g. [78, 74, 37]. Such models are used in wind speed or power forecasting in e.g. [14, 75, 33]. The principle of wavelet decomposition and AR time series modeling has been presented in Section 4.3. As the results presented in the following Section 7 show that generally the LS-SVM model yields the most accurate results, we hereby describe its structure in more detail following [30, 22, 69].

The roots of LS-SVM lie in standard SVM modeling technique introduced by Vapnik [82]. A support vector machine constructs a set of hyperplanes in a high-dimensional space. The hyperplanes act as separators which can be used in predicting in which class a new data point belongs. An illustration on the principle of a SVM is presented in Figure 28.

The SVM solutions are characterized by convex optimization problems and a few tuning parameters. It can nevertheless turn out to be quite challenging to solve

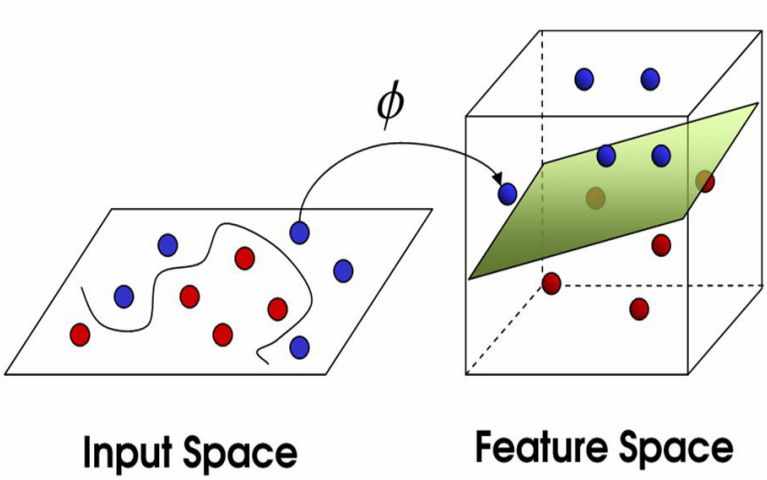

Figure 28: The mapping of inputs into a feature space in a support vector machine. Source: http://www.imtech.res.in/raghava/

the SVM problem, as in primal space the problem involves inequality constraints and in dual space the solution must be calculated using *quadratic programming* techniques. These inconveniences can be overcome using a reformulation of the original algorithm called LS-SVM, with which the problems can be expressed as linear *Karush-Kuhn-Tucker* systems. The only drawback is that as in the traditional SVM formulation the solution is sparse in the sense that only certain points used in training contribute to the solution, in LS-SVM every point has significance. On the other hand, this is the case in more common neural network architectures as well. There are furthermore LS-SVM pruning techniques which can be used to obtain a degree of sparsity.

Again, the aim of LS-SVM modeling is to estimate a mapping f between the explanatory and the output variables according to (37). The LS-SVM modeling approach does this by building a nonlinear representation of the original inputs in a high-dimensional feature space using positive-definite kernel functions.

Let us rewrite (37) as

$$p_t = \boldsymbol{\mu}^T \varphi(\Phi_t) + b + e_t \tag{69}$$

where $\varphi(\cdot)$ is a nonlinear feature mapping transforming the original inputs Φ_t to a high-dimensional (or even infinite-dimensional) vector $\varphi(\Phi_t)$. Then in order to obtain p_t, weights $\boldsymbol{\mu}$ and a bias term b are applied. Now, we consider a constrained *ridge regression* optimization problem of the form

$$\begin{aligned} \min_{\boldsymbol{\mu},b,e_t} \mathscr{J}(\boldsymbol{\mu}, e_t) &= \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\mu} + \frac{\wp}{2}\sum_{t=1}^{N} e_t^2 \\ \text{s.t. } p_t &= \boldsymbol{\mu}^T \varphi(\Phi_t) + b + e_t \end{aligned} \tag{70}$$

where $\wp$ is called a *regularization constant*, which can be understood as a measure of trade-off between the smoothness and accuracy characteristics of the fit. In order to solve (70), a standard Lagrangian function with Lagrange multipliers α_t is constructed:

$$\mathscr{L}(\boldsymbol{\mu}, b, e_t, \alpha_t) = \mathscr{J}(\boldsymbol{\mu}, e_t) - \sum_{t=1}^{N} \alpha_t \left[\boldsymbol{\mu}^T \varphi(\Phi_t) + b + e_t - p_t\right] \tag{71}$$

Now the necessary optimality conditions $\nabla\mathscr{L}(\boldsymbol{\mu}, b, e_t, \alpha_t) = \mathbf{0}$ can be summarized in a matrix form as

$$\left(\begin{array}{c|c} 0 & \mathbf{1}_N^T \\ \hline \mathbf{1}_N & \boldsymbol{\Omega} + \wp^{-1}\boldsymbol{I}_N \end{array}\right)\left(\begin{array}{c} b \\ \hline \boldsymbol{\alpha} \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline \boldsymbol{p} \end{array}\right) \tag{72}$$

where $\mathbf{1}_N = [1 \dots 1]^T$, $\boldsymbol{I}_N$ = N x N identity matrix, $\boldsymbol{p} = [p_1 \dots p_N]$ and $\boldsymbol{\alpha} = [\alpha_1 \dots \alpha_N]$. According to *Mercer's theorem*, there exists an expansion so that $\Omega_{t_1 t_2} = \varphi(\Phi_{t_1})^T \varphi(\Phi_{t_2}) = K(\Phi_{t_1}, \Phi_{t_2})$. Note that using such notation, the mapping function $\varphi(\cdot)$ never needs to be explicitly calculated, as in (72) they only appear in dot products which are replaced by *kernels*. A common choice for the kernel $K(\cdot, \cdot)$ is a RBF (Gaussian) kernel of the form $K(a, b) = \exp(-||a - b||_2^2/\ell)$ where ℓ is called the *kernel bandwidth*. By using such a kernel the feature space can be shown to be a Hilbert space of infinite dimension [77].

From (72) one obtains values for α and b. Using them, the resulting LS-SVM function estimate in a new point ζ can be written as

$$\hat{f}(\zeta; b, \alpha) = \sum_{t=1}^{N} \alpha_t K(\zeta, \Phi_t) + b \tag{73}$$

Note that the LS-SVM structure is dependent on two tuning *hyperparameters*: the regularization constant $\wp$ and the kernel bandwidth ℓ. They can be optimized by using a cross-validation method, where data is first divided into disjunct sets and then the different sets are used in estimating the performance of models trained with other sets. By measuring the performance of each candidate model with different hyperparameters, an estimate of optimal hyperparameters can be obtained. What is also noteworthy that in (LS-)SVM models the solution of optimization problem is global, which is not the case in traditional neural network (FF-BP, RBF) architectures.

For making uncertainty estimations, we adopt the quantile regression technique presented in Section 4.3.5. The mapping function in (61) is chosen to be a simple linear one:

$$f_j^{(QR)}(x_j; \omega) = \beta_j(\omega) x_j \tag{74}$$

The AR(X)-model for every wavelet-decomposed component of the original measured wind speed signal is identified recursively using a forgetting factor of 0.99. The lags are chosen according to the AIC criterion (44) calculated on the training set of the data. This means that the lag parameters may be different at different instants. The training set is based on historical winds up to forecasting instant starting from 1000 hours before that.

There are consequently as many models as the level J of the wavelet decomposition imposes. However, only for the approximation part the original NWP is used as an external input, and for the details the models are pure AR models without any external variables. The individual models $(\mathrm{AR}(p^{(A)}, \upsilon), \mathrm{AR}(p^{(D)}))$ and the final new prediction $\tilde{\hat{w}}$ may thus be written as

$$\begin{aligned} w_t^{(A)} &= \sum_{i=1}^{p^{(A)}} \varphi_i w_{t-i}^{(A)} + \sum_{k=0}^{\upsilon} \theta_k \hat{w}_{t-k} + \epsilon_t \\ w_t^{(D_j)} &= \sum_{i=1}^{p^{(D_j)}} \varphi_i w_{t-i}^{(D_j)} + \epsilon_t \;\forall j \in J \\ \tilde{\hat{w}}_t &= w_t^{(A)} + \sum_{J} w_t^{(D_j)} \end{aligned} \tag{75}$$

6.3 Implementation

For testing purposes, all the models are implemented in MATLAB environment. The needed add-ons include the Fuzzy Logic, Neural Network, Statistics, System Identification and Wavelet toolboxes. A freely available toolbox [63] was used for LS-SVM model implementation. For making quantile estimations, we used another freely available *interior point method* based function [35]. Later on, the models are going to be implemented on a SCADA system run by the FWPP but at this stage such implementation is not considered. Using data described by Table 3 such later implementation is however possible, as long as the SCADA system provides full capabilities in programming the needed calculations.

We estimate three (3) different sets of forecasting models. The first group, which is hereafter named as Model Ensemble 1 (ME1), consists of models described in Table 4 apart from the wind speed model on the last line. Those models are trained with exactly the inputs shown in Table 4. All those models are one-step forecasters, which means that they produce a prediction of one step ahead to prediction instant. In order to predict further, the models are used iteratively so that autoregressive inputs are replaced by their corresponding predictions.

The second group, ME2, contains the same models as ME1. The difference is in their training: it is carried out by using inputs modified with replacing the wind and wind direction forecasts by their real measured values. In this way, the ME1 group incorporates a sort of internal NWP bias correction as it tries to estimate the

relationship of predicted weather information and measured output. On the other hand, in ME2 we wanted to form a relationship between the real measured values of wind speed and production as accurately as possible, without any anticipatory wind speed bias correction whatsoever. This is done because with the last wavelet-AR(X) model in Table 4 we want to form another forecast of wind, possibly trying to account for recent biases in NWPs. Such new wind speed forecasts are fed into ME2 with the other values remaining intact.

The third group, ME3, contains again the same models as in Table 4. Now, however, we want to test making directly k-step ($k > 1$) forecasts, so that the models are trained to predict the production k steps ahead to prediction instant. As there are a lot of different interesting forecasting steps and in theory the inputs could be of different nature in every case, in this thesis we restrict the testing to making 3 and 6-step forecasts with a similar (but optimized) input structure.

Figure 29 clarifies the use of the models described above:

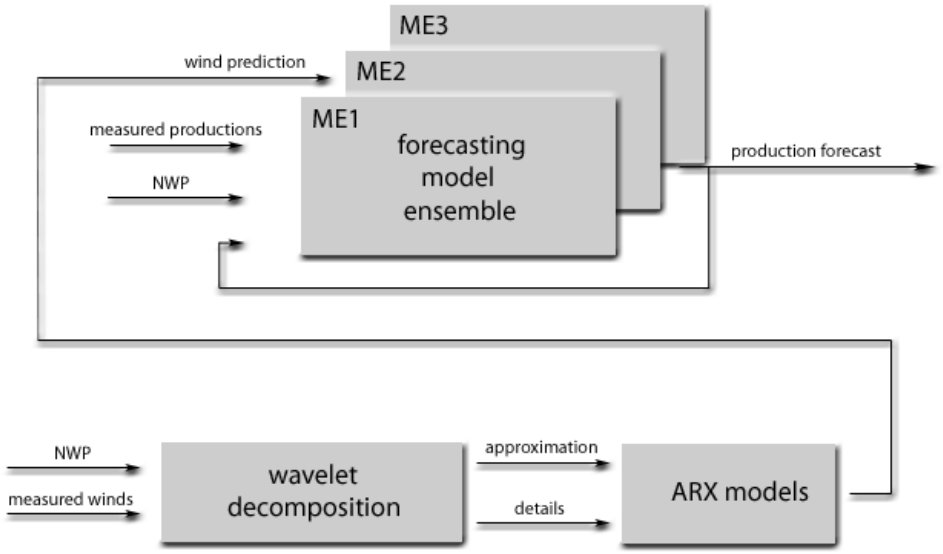

Figure 29: Chart of forecasting models and their inputs

Mathematically we may express the model ensembles as follows:

$$\begin{aligned} \boldsymbol{p}_{t+1|t}^{(\mathrm{ME1})} &= \boldsymbol{f}(\Phi_t^{(\mathrm{ME1})}) \\ \boldsymbol{p}_{t+1|t}^{(\mathrm{ME2})} &= \boldsymbol{g}(\Phi_t^{(\mathrm{ME2})}) \\ \boldsymbol{p}_{t+k|t}^{(\mathrm{ME3})} &= \boldsymbol{h}(\Phi_t^{(\mathrm{ME3})}) \\ \tilde{\hat{w}}_{t+1|t} &= \boldsymbol{c}(\boldsymbol{w}_t) \end{aligned} \tag{76}$$

where the training sets are defined as

$$\begin{aligned}
\Phi_t^{(\mathrm{ME1,train})} &= \left[p_t, \ldots, p_{t-l_p}, \hat{w}_{t+1}, \ldots, \hat{w}_{t+l_w}, \cos(\hat{d}_{t+1}), \sin(\hat{d}_{t+1})\right] \\
\Phi_t^{(\mathrm{ME2,train})} &= \left[p_t, w_{t+1}, \cos(d_{t+1}), \sin(d_{t+1})\right] \\
\Phi_t^{(\mathrm{ME3,train})} &= \Phi_t^{(\mathrm{ME1,train})}
\end{aligned} \tag{77}$$

whereas the information sets used in the actual testing (validation sets) are written as

$$\begin{aligned}
\Phi_t^{(\mathrm{ME1})} &= \Phi_t^{(\mathrm{ME1,train})} \\
\Phi_t^{(\mathrm{ME2})} &= \left[p_t, \tilde{\hat{w}}_{t+1}, \cos(\hat{d}_{t+1}), \sin(\hat{d}_{t+1})\right] \\
\Phi_t^{(\mathrm{ME3})} &= \Phi_t^{(\mathrm{ME3,train})}
\end{aligned} \tag{78}$$

Note that the issuing instants have been dropped in (77) and (78) for simplicity.

7 Results

In this Section, we investigate the forecasting accuracy of the models presented in Section 6. The results are compared with the current procedure in order to justify the need for building an advanced model for reducing the forecasting inaccuracy. Also the accuracy gain transformed into cash flow increase is discussed. Furthermore, we estimate the level of remaining uncertainty when using an advanced forecasting model.

7.1 Performance of the implemented models vis-à-vis reference model

The performance tests are carried out using the following procedure. Firstly, we use a training set of 1520 hours (approximately 2 months, see Figure 30), which is a subset of the whole data available in order to reduce the computational load. It is also important not to use too long a training set in order to account for significant seasonal variations in wind speed/production data. The training set is used to train the models described in Section 6.2. The training procedure is repeated several times with different numbers of power output lags and future NWPs. This procedure aiming at selecting the best combination of explanatory variables is done only for LS-SVM model as the time did not allow us to do the selection individually for all other models. It is rather believed that the combination which is performant with LS-SVM model is acceptably performant for other models as well, although naturally there is no general guarantee on that. Note also that the performance of LS-SVM is a function of its two hyperparameters. They are at first fixed and subsequently optimized.

The performance of different models is assessed using a RMSE-metric (39) over a validation set of 470 hours, which corresponds to approximately 20 days. The results of input selection based on varying the number of power output lags and future NPWs are presented in Figures 31 and 32. They present the RMS error calculated over the validation set using different inputs, while preserving another one of them at constant value 1. Based on these results, it was deemed that an optimal input would be one where 4 lags of historical power and 3 future steps of NWP information are employed. Note that different look-ahead steps of NWP concerning the wind direction are not tested here, as its explanation power is much less important than the magnitude of wind itself. The optimal ME1-input used hereafter is thus written as

$$\Phi_t^{*(\mathrm{ME1})} = \left[p_t, p_{t-1}, p_{t-2}, p_{t-3}, \hat{w}_{t+1}, \hat{w}_{t+2}, \hat{w}_{t+3}, \cos(\hat{d}_{t+1}), \sin(\hat{d}_{t+1})\right] \tag{79}$$

For the ME2 set it was decided to use a simple (1 power lag, 1 NWP) approach as the quality of competing wind predictions $\tilde{\hat{w}}$ would anyway decrease very quickly after a few look-ahead times. It would also be difficult to account for $\tilde{\hat{w}}$-bias as it

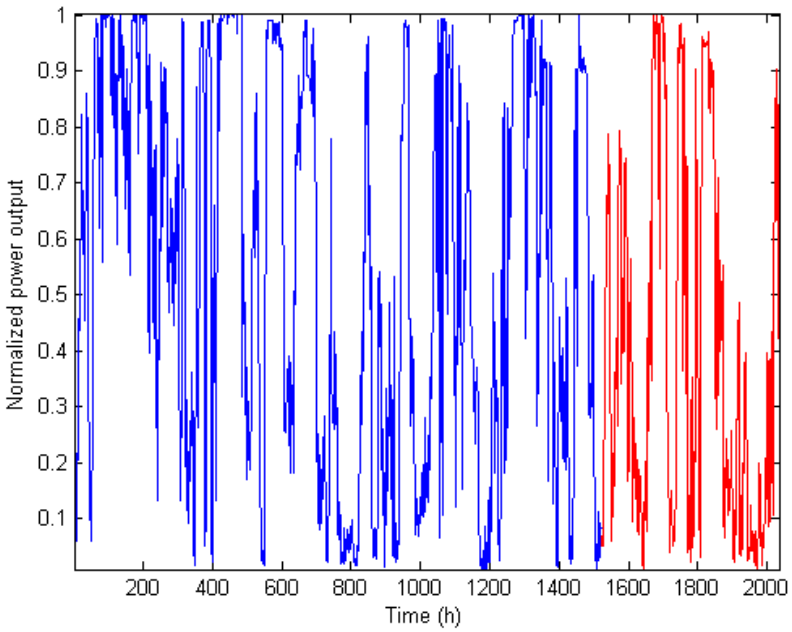

Figure 30: Measured production in the training (blue) and validation (red) sets

probably is very unsystematic. Furthermore, the nature of ME2 is to make only very short-term predictions so that a direct wind-production -relationship is preferred for simplicity. We can now write

$$\Phi_t^{*(\mathrm{ME2})} = \left[p_t, \tilde{\hat{w}}_{t+1}, \cos(\hat{d}_{t+1}), \sin(\hat{d}_{t+1})\right] \tag{80}$$

In what comes to the ME3 set, it was decided to make similar tests than with ME1 in order to find the optimal input for both 3 and 6-step forecasters. The results of those tests are presented in Figure 33. However, the wind forecast step was fixed to the corresponding forecasting instant in order to economize on the number of tests to be made. Based on the tests, it seems that a simple 1-step lagged version of autoregressive input would be optimal and parsimonious for both the 3- and 6-step ahead predictions. We thus have

$$\Phi_t^{*(\mathrm{ME3})} = \left[p_t, \hat{w}_{t+1}, \cos(\hat{d}_{t+1}), \sin(\hat{d}_{t+1})\right] \tag{81}$$

The reader may note that the performance of all models with different inputs seems to get better when passing the limit of 24 hours in forecasting look-ahead time. This is explained by two factors: firstly, there naturally is some variation on the performance as a function of the validation set. By choosing another set (which could be longer for instance) we might get different results. Another more important reason probably pertaining in every test is that in our data the maximum age of a NWP is 24 hours, as they were originally stored in a vector comprising a single value for every single hour throughout the dataset. In reality, the lifetime of a NWP is

38 hours, as explained in Section 5.3. However the last $38 - 24 = 14$ hours are not stored in the data system but are being replaced by new ones arriving every 24 hours. Thus the results for making a spot market forecast (which requires forecasting at horizons ranging from 13 to 36 hours) are slightly discussable after 24 hours. It is nevertheless believed that based on findings in Figure 22, the relative accuracy of NWPs stays quite constant after the few first look-ahead times. With this kept in mind it can be argued that probably the performance of models in Figures 31 and 32 stays approximately constant after 24 hours as well.

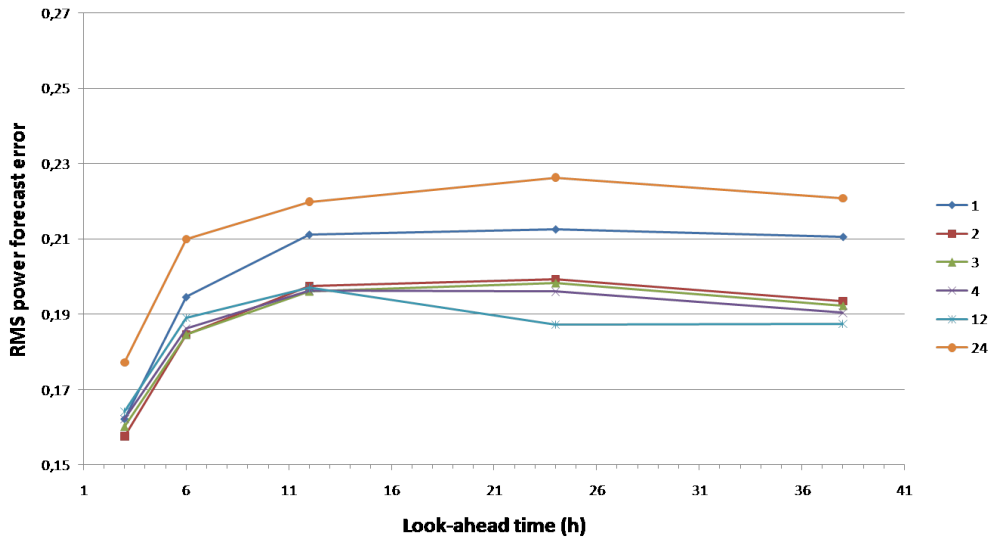

Figure 31: LS-SVM(ME1) model performance with different power lags l_p, keeping NWP step at 1

With the input now fixed, we can compare the behavior and performance of all individual models in sets ME1 and ME2. Let us however first apply a simple input using 1 power lag and 1 NWP in order to identify the general performance of different model candidates. This is done for the set ME1 and the results are presented in the left part of Figure 35. The individual characteristics of the testing candidate models are presented in Table 5. We can see that the performance of ANFIS is rather mediocre, so we can drop it from both sets. Another reason for this is that when trying to apply the identified optimal input in the ANFIS model its results were even worse.

Table 5: ME1 candidate model parameters and characteristics

Model	Characteristics
FF-BP	5 neurons, 2 layers, tanh activation function
RBF	15 neurons, 2 layers, RBF spread 10
ANFIS	Sugeno-type grid partition, 2 Gaussian MFs
LS-SVM	$\wp = \ell = 10$

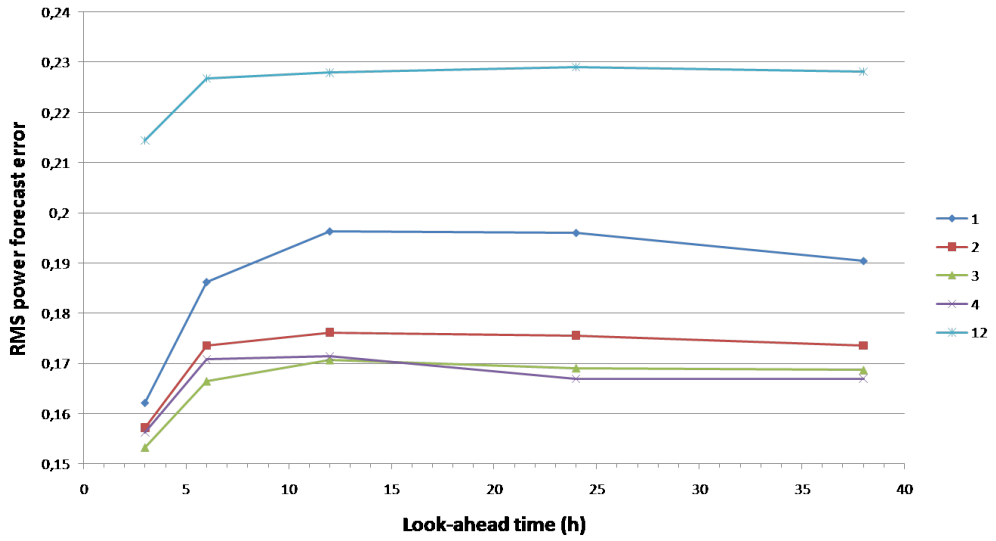

Figure 32: LS-SVM(ME1) model performance with different steps of NWPs (l_χ), keeping power lag at 1

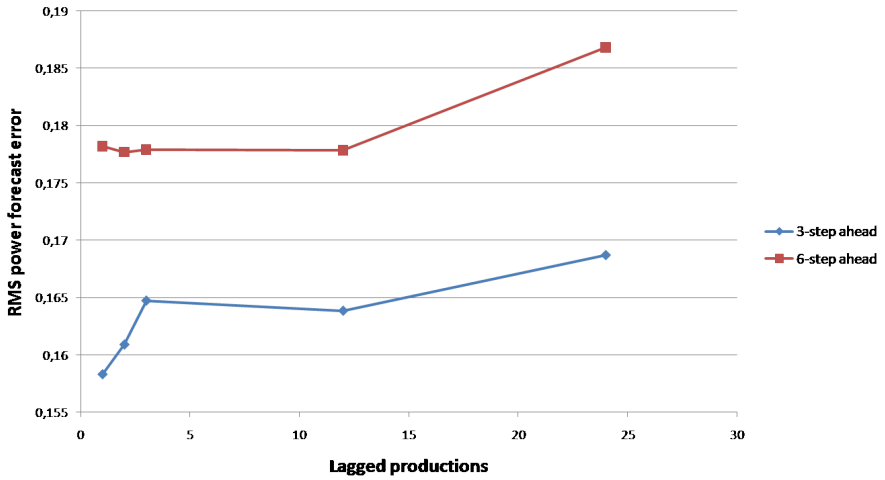

Figure 33: LS-SVM(ME3) model performance with different power lags l_p when making 3- and 6-step ahead predictions

Now, we compare the performance of the resting four models using the optimal input. An illustration of ME1-forecasting is presented in Figure 34 and the results are presented in the right part of Figure 35. Note that Figure 34 also presents ensemble forecasting based on optimal weights calculated using (54) as well as naïve equal weights. We did not however find any real improvement with such an approach so it is abandoned in the further discussion. Some possible reasons for that are revealed in Section 4.3.4.

Looking at Figure 35 it is clear that with the available validation data, the most accurate forecaster is the LS-SVM model. It seems to beat all the other models

at every look-ahead time. The reasons for this are probably various and at least clearly not very evident; however there are some aspects that may play a role in the superior performance. The optimization solution of LS-SVM is unique compared to RBF and FF-BP models, where the solutions are typically only locally optimal. The worst performance of FF-BP network is explained by its global approximator characteristic. The RBF network is a local function approximator, but it is also dependent on its weights in a sense that they may not prove to be globally optimal.

We also show the effect of forecasting look-ahead time for 1, 6, 12 and 24 hour horizons in Figure 36, which shows the forecasting results by successively applying a k-step forecast and plotting the result against the measured data. Clearly, the larger the forecasting horizon is, the smoother and the more inaccurate the forecast becomes. It also has to more and more rely on the NWP input as its driver.

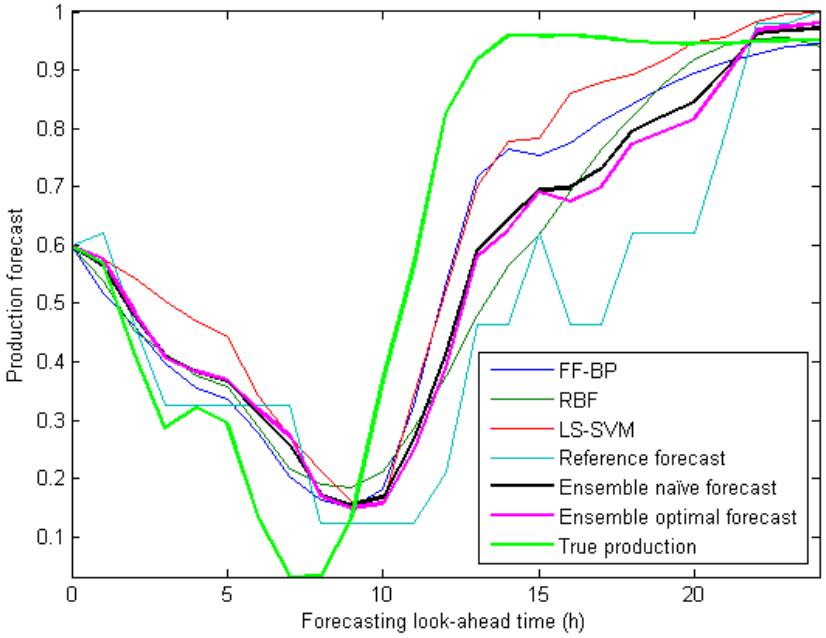

Figure 34: ME1-forecasting illustration using the optimal input

The RMSE-metric based improvement results according to (40) are presented in Figure 37. Comparing the performance of the LS-SVM model (in ME1) to the reference model, we can see that the effort of building an advanced model pays well off: even at the largest forecasting look-ahead time (24 hours, after which the results are not reliable due to reasons explained above), the relative improvement compared to the original reference model is over 25%. The same improvement can be seen also from Figure 38, where a 36-hour forecast has been made every hour over the validation data and the average RMS error over those forecast steps has been calculated for the LS-SVM(ME1) and the reference model. The advanced model clearly beats the reference almost every time a forecast is made.

If one wants to investigate whether the average errors of prediction are mainly

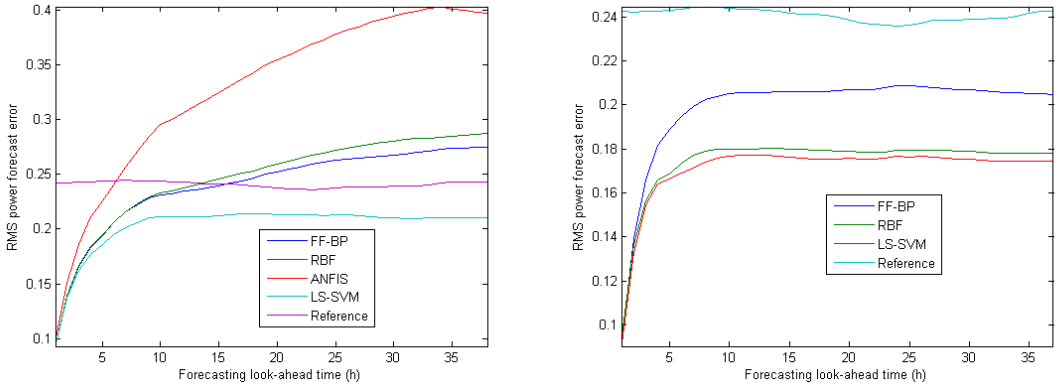

Figure 35: ME1 performance using a simple input and the optimal input

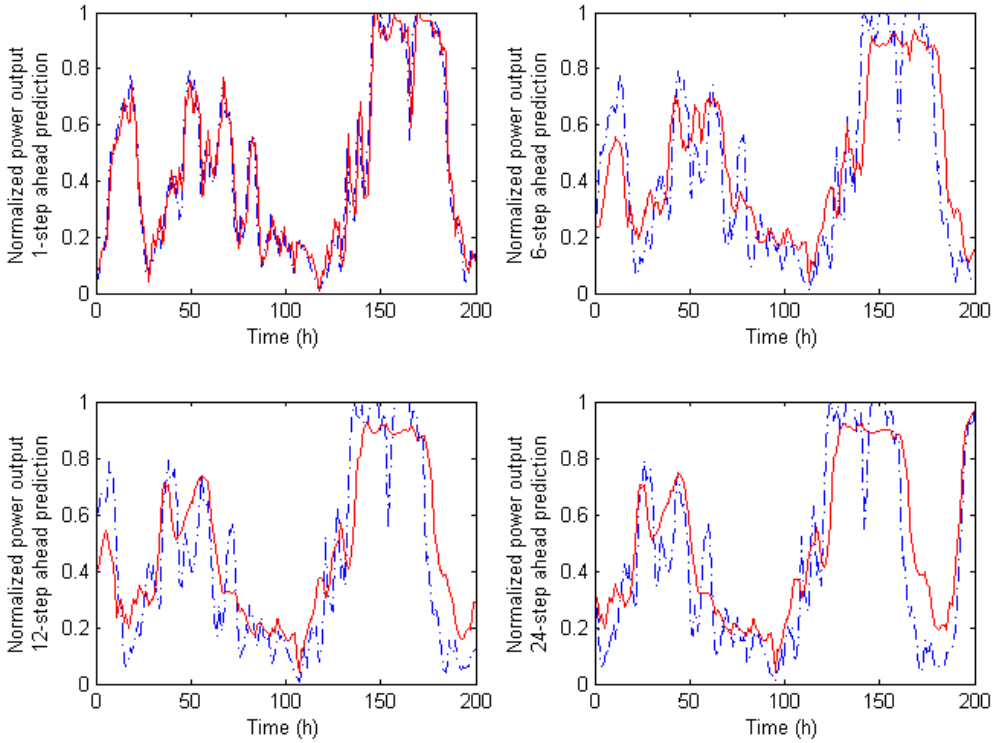

Figure 36: The effect of forecasting look-ahead time on the forecasting performance. Blue (dashed) = measured, red (solid) = predicted production

amplitude or phase errors by nature, the decomposition in (42) may be applied. Figure 39 presents the results of such decomposition as a function of forecasting look-ahead time. What is notable is that clearly the amplitude errors reflected by the bias and sdbias terms in (42) are almost zero, as the components almost annul each other. The average error is hence primarily a function of phase shift errors. They, on the other hand, are mainly related to NWP phase errors, as shown in [43],

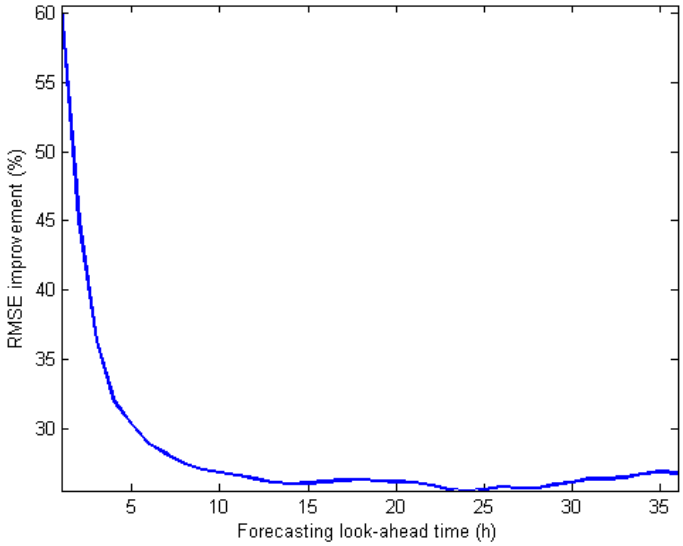

Figure 37: RMSE improvement of LS-SVM model compared to the reference power curve model

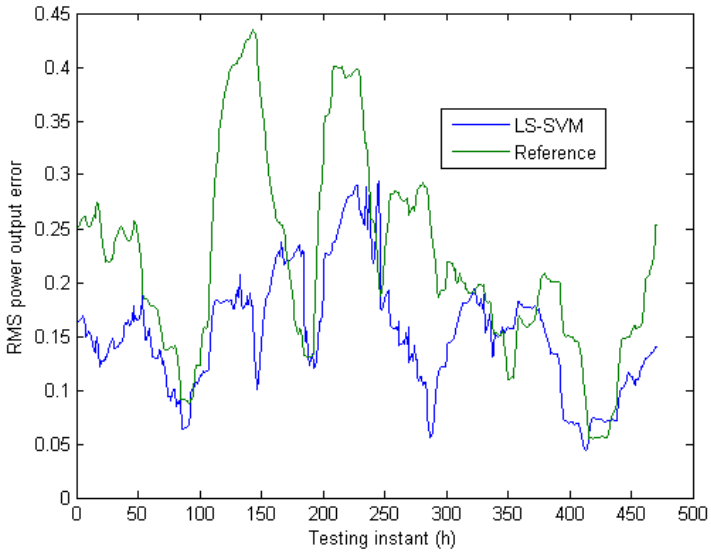

Figure 38: Average 36-hour forecast error using LS-SVM and reference models

so that the initial prediction errors of wind speed are directly transferred to power prediction errors. What this means is that the quality of NWP is clearly of essential value.

Relying on the fact that the LS-SVM would be the most accurate model, we can now train and validate the ME2 ensemble consisting only of a LS-SVM model. The training is based on principles presented in Section 6.2.

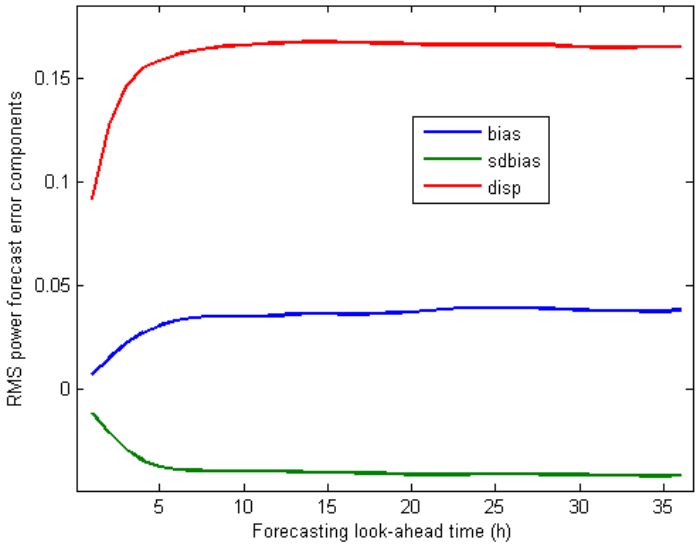

Figure 39: Decomposition of RMS forecast error according to look-ahead time

As a first test, we want to show that using NWP-based training (ME1) is more efficient than training the same model with realized winds and then changing the input to contain NWPs. This can be tested by comparing the performance of ME1 and ME2 with same inputs. The results, which show a superior performance of ME1, are shown in Figure 40. Note also that the reference method here beats the advanced ME2 model, which simply reflects the fact that the reference power curve is seemingly not a one-to-one turbine manufacturer power curve (although based on it) but tries also to account for average NWP errors.

Now, if we change the original NWPs of ME2 (validation set) to our own wavelet-AR(X)-based predictions, we obtain a quite different result. An illustration on forming the wavelet-AR(X) wind speed forecast is presented in Figure 41. A comparison of ME1 and ME2 model performances with forecasting look-ahead times from 1 to 6 hours is presented in Table 6. It can be seen that the new homemade wind forecast is slightly more performant in short look-ahead times. The reason for this is that probably there are a lot of situations in which the NWP is strongly biased, so that the effort of making another wind forecast pays off. However in longer look-ahead times one clearly must stick to the original NWP information. Furthermore, the extra effort in making another wind forecast with respect to the improvement may be subject to debate.

For deciding whether to employ a 1- or multistep forecaster, we compare the performance of ME1 to the one of ME3 using the LS-SVM model with respective optimal inputs as their only member. The results for making 3- and 6-step ahead predictions for both models are shown in Table 7. Based on these tests, it is shown that at least for very short-term prediction, it is preferable to iteratively use a 1-step forecaster rather than a pure k-step forecaster. The difference of performances is not however

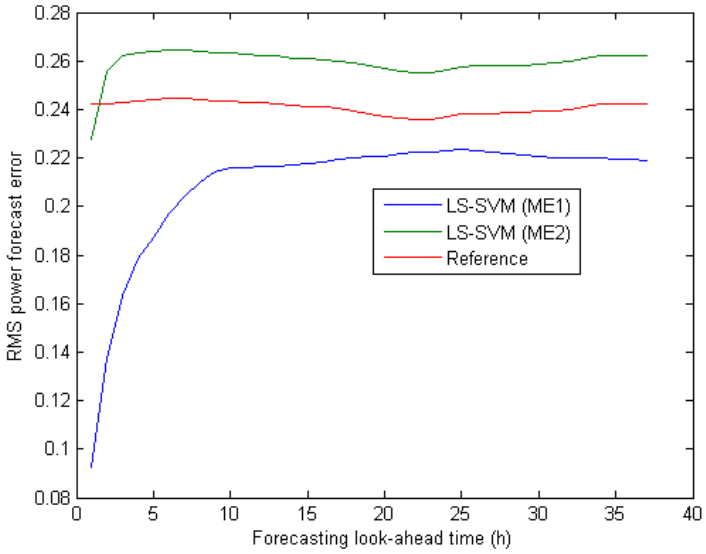

Figure 40: Comparison of LS-SVM models trained with real (ME2) or NWP-based (ME1) winds

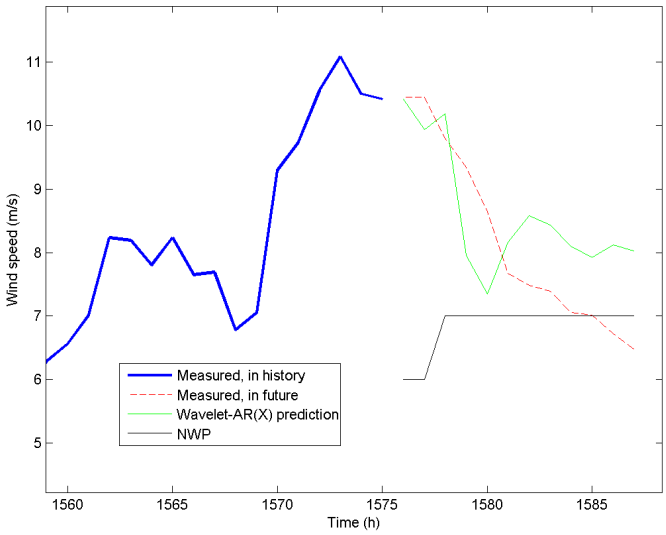

Figure 41: Wavelet-AR(X) model forecasting illustration

very big and is definitely a function of the used validation data. The reasons for this difference remain somewhat unclear, although probably the most intuitive one is the best one: the error of iterative forecasting does not grow at the same speed as the error of direct modeling of information which is further apart from the prediction basetime. Such models in a way neglect possible outcomes of production between the basetime and the desired prediction instant. As discussed earlier, in the literature there are recommendations and implementations of both approaches.

Table 6: ME1 and ME2 comparison on very-short term forecasting accuracy

Look-ahead time	RMSE(ME1)	RMSE(ME2)
1	0.0905	0.1099
2	0.1331	0.1249
3	0.1548	0.1426
4	0.1645	0.1609
5	0.1664	0.1787
6	0.1691	0.2042

Table 7: ME1 and ME3 RMS errors in 3- and 6-step ahead forecasts

Model	3-step	6-step
ME1	0.1546	0.1692
ME3	0.1584	0.1782

Finally, we would like to investigate the effect of model tuning and pruning on its performance. On the other hand, it is investigated whether a robust training of the model would be appropriate in order to account for possible outliers of non-Gaussian noise. Such training can be carried out by reducing the impact of support vectors which lead to errors [63]. We employ the earlier obtained optimal input first to a LS-SVM(ME1) model with randomly chosen ($\wp = 10, \ell = 10$) hyperparameters. Then we tune its parameters with the same inputs and record the performance with the same validation data. As discussed, the LS-SVM structure is not sparse so that every training point corresponds to the solution. In theory, if trained further and further the structure grows all the time and so the model implementation becomes somewhat complex, possibly modeling outliers as well. There are methods for removing the least relevant support vectors in order to obtain a degree of sparsity. We test this by fixing the optimal input and hyperparameters and test the performance of the pruned model compared to the original full one. The accuracy results in the form of RMSE metric are presented in Figure 42.

From Figure 42 it can be read that the optimization of the hyperparameters does not pay off: as it concentrates only on the 1-step forecasting (and definitely slightly beats the original version in that) the performance in further time steps gets worse than with the original model. Consequently, in the future care must be taken in order to choose efficient hyperparameters that work well if an iterative 1-step forecaster is going to be used. However, the pruning procedure can be seen to be quite innocent in reducing the average performance. This is in agreement with the general idea of every point not having a very significant impact on the solution. The same can be seen when robust training is performed: such a model beats the rest at almost every look-ahead time. Evidently in the data there has been outliers or non-Gaussian noise, whose negative performance effect can be dampened by applying robust estimation methods.

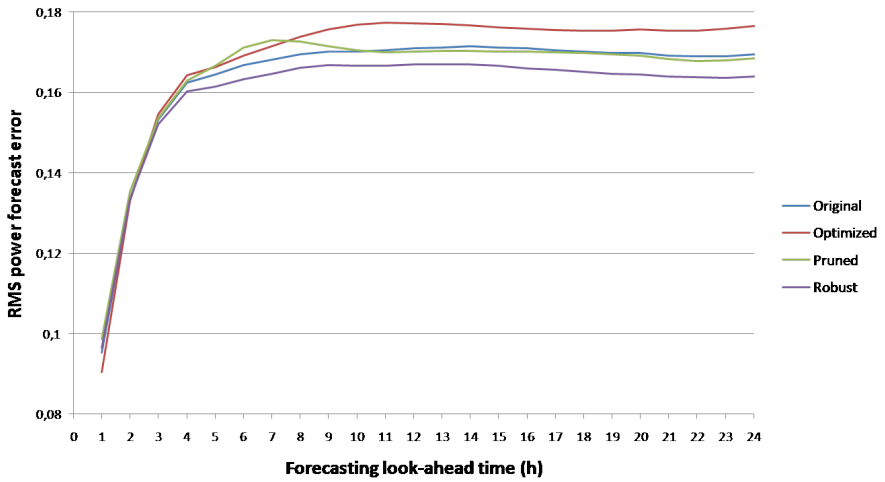

Figure 42: The effect of tuning, pruning and robust training on LS-SVM performance

7.2 Effect on net revenue

We briefly investigate the effect of using an advanced wind power forecasting system on the net revenue (16) it obtains while participating in the electricity markets. The advanced model used in testing is the LS-SVM(ME1) described in Section 7.1. The testing data here consists of the same validation set as before, for which first spot market forecasts are made and then the regulation costs calculated. We also consider the effect of participating in the intraday markets on the net revenue. The limitations of intraday markets must be borne in mind: in reality, there is no single price as in the spot markets. In our data, the average hourly price is used and it is assumed that the intraday markets are sufficiently liquid, i.e. there always exists an intraday bid corresponding to the volume WPP wants at the average price marked in the data.

The testing procedure is the following: firstly, we calculate revenues using the reference and the advanced model. At this stage, no intraday market participation is considered. Then, the same calculation is remade using the assumption that the WPP can make short-term forecasts and use their information in intraday trading. The short-term forecasts are made 6 and 3 hours ahead, so that the significance of look-ahead time can be distinguished. In order to remove the effect of predicting the correct number of available turbines in the future, all the calculations are made for normalized data, so that they represent cash flow one would obtain using a single 1 MW wind turbine.

The overall net revenue calculated as a *daily mean* over the testing data as well as the relative *improvement with respect to the reference model* with only day-ahead trades are presented in Table 8. In the Table, S refers to day-ahead trades (forecast horizon 13–36 hours), I to intraday trades and PP to perfect power forecasts.

From the results of Table 8 the reader may note that firstly, the effect of making intraday trades with the current reference forecaster is quite limited. The reason for

Table 8: Net revenue with reference and advanced forecasting models

Reference		S	S+ 6-h I	S+ 3-h I	PP
	spot trades	233.91 €	233.91 €	233.91 €	343.36 €
	intraday trades		0.59 €	0.59 €	
	balance energy trades	71.06 €	70.87 €	70.87 €	
	net revenue	304.97 €	305.37 €	305.37 €	343.36 €
	improvement	0.00%	0.13%	0.13%	11.18%
Advanced					
	spot trades	373.01 €	373.01 €	373.01 €	343.36 €
	intraday trades		-1.86 €	-13.24 €	
	balance energy trades	-47.96 €	-44.37 €	-29.44 €	
	net revenue	325.06 €	326.79 €	330.34 €	343.36 €
	improvement	6.18%	6.68%	7.68%	11.18%

this is obvious: as seen earlier (Figure 22) only very recent NWPs provide information that is statistically more accurate than older ones.

Furthermore, one notes that when using an advanced forecaster system and a 3-step intraday forecast with corresponding intraday trades, the daily mean improvement of net revenue is roughly 7.7%. At first, it may seem to be quite mediocre, as the overall accuracy improvement gained by using an advanced forecaster (see Figure 37) is much more. There are two reasons for this: as shown in 3.4, it generally may not be optimal to bid the predicted amount (if the prediction is not completely accurate all the time) without considering the prediction uncertainty. Secondly, the electricity prices have naturally an effect on the financial performance: they either dampen or fortify the prediction inaccuracies. In this case it seems that they are not extremely penalizing. The speciality of Finnish balancing markets of not penalizing producers helping to balance deviations means that there are several occasions under which regulation does not cost any extra compared to a situation where production would have been predicted correctly. In any case, any daily net revenue improvement cumulates quickly to big increases in e.g. monthly income, so that the use of advanced techniques is strongly motivated.

7.3 Estimating the uncertainty

Even if the advanced models produce forecasts which are more accurate than the current power curve model, there are still significant errors and thus uncertainty related to power forecasts. For estimating the level of this uncertainty, we apply the techniques presented in Section 4.3.5. More specifically, we want to estimate the cumulative distribution function of the future power output based on power forecasts. This can be done using quantile regression.

We use the LS-SVM(ME1)-model with optimal inputs and estimate the 5%, 25%, 50%, 75% and 95% ($n = 5$) probabilities of really producing less than the predicted

amount. This must be done separately for every forecasting look-ahead time k, so that we finally obtain $k \cdot n$ parameters β from the quantile regression. An illustrative example of augmenting point forecasts with uncertainty information in the form of probability quantiles is presented in Figure 43. The estimated quantile regression coefficients $[\beta_0, \beta_1]^T$ are expressed in Table 9.

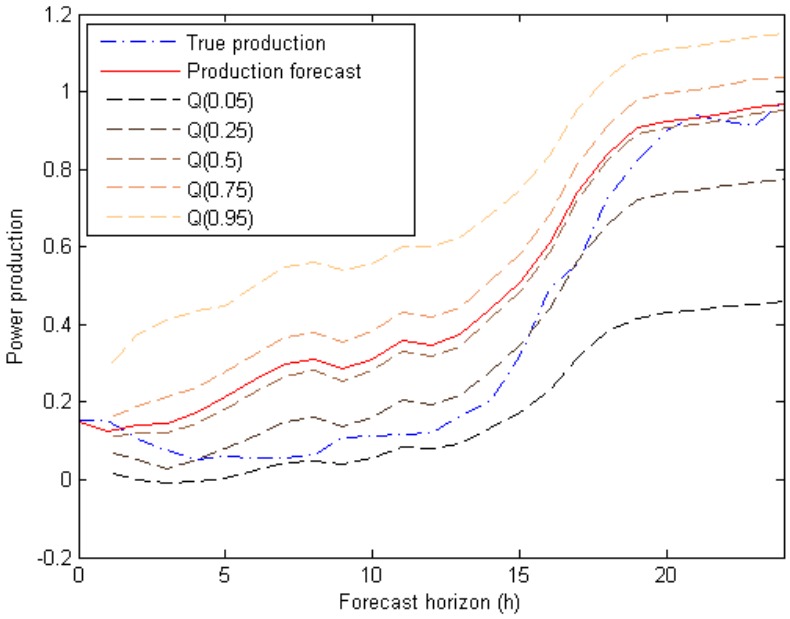

Figure 43: Production forecast and its estimated quantiles

From Figure 44 we can see that clearly the uncertainty increases with the forecasting look-ahead time. In the Figure it has been predicted that in 1, 6, 12 and 24 hours the production will be 0.5. Then corresponding 5%, 25%, 50%, 75% and 95% quantiles representing the amount under which the real production will fall given the probability are shown. What is notable also is that the forecast does not correspond to the 50% quantile. This is in agreement with the fact that the forecasting model minimizes the sum of squared errors of the fit, whilst the 50% quantile represents the median of the measured production. These two concepts can certainly yield different results.

However, in order to use this uncertainty information operationally, the reliability of obtained quantiles must be verified. This can be done by matching the observed production frequencies, given a range of forecasts, with the nominal quantile probabilities. For this purpose, we have divided the production forecasts into four parts, ranging with 25% increases from zero to full production. The production forecasts are 6-step ahead forecasts. The results are presented in Table 10.

From the results we can see, that although there are some observed values that are in good agreement with the nominal probabilities, there are some points where significant deviations occur. The reasons for this are probably three-fold: a larger number

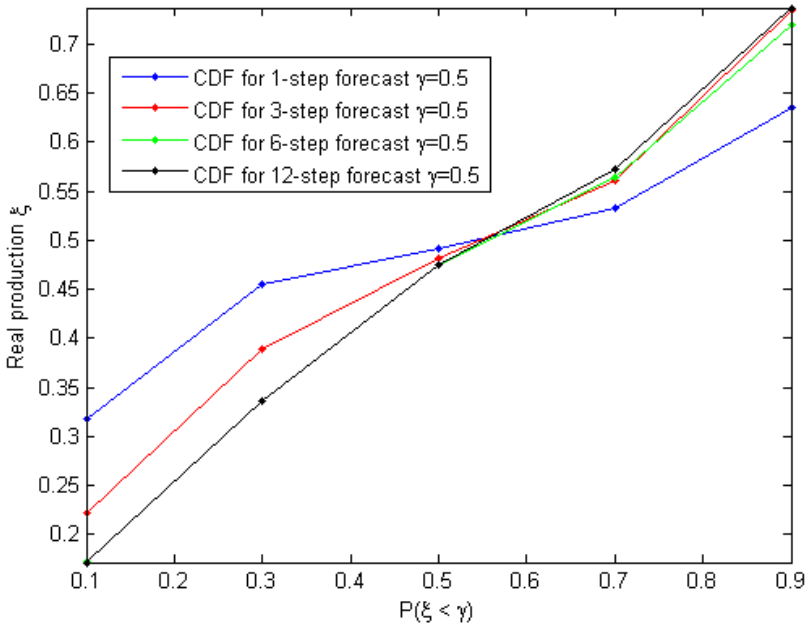

Figure 44: Quantiles of predicted production at different forecasting look-ahead times

Table 9: Estimated quantile regression coefficients

Look-ahead time (h)	$\boldsymbol{\beta}(0.05)$	$\boldsymbol{\beta}(0.25)$	$\boldsymbol{\beta}(0.50)$	$\boldsymbol{\beta}(0.75)$	$\boldsymbol{\beta}(0.95)$
1	-0.0801	-0.0590	-0.0218	0.0323	0.1664
	0.7961	1.0273	1.0252	1.0015	0.9370
6	-0.1362	-0.1246	-0.0396	0.0637	0.2632
	0.6164	0.9221	1.0287	1.0010	0.9133
12	-0.1316	-0.1309	-0.0376	0.0693	0.2878
	0.6063	0.9335	1.0239	1.0062	0.8987
24	-0.1401	-0.1383	-0.0396	0.0661	0.2759
	0.6191	0.9444	1.0245	1.0057	0.9043

of samples (than the current length of validation set) should have been used in order to attach the probabilities of production given a certain forecast more realiably. Also the number of observed productions given a forecast range was rather limited in the validation set, so that there may be some selectional bias. A simple linear relationship (58) used in quantile estimation might neither be sufficient to describe the relationship of predicted and measured production in all cases. The deviations of nominal and observed production probabilities conditioned on predictions are presented in Figure 45.

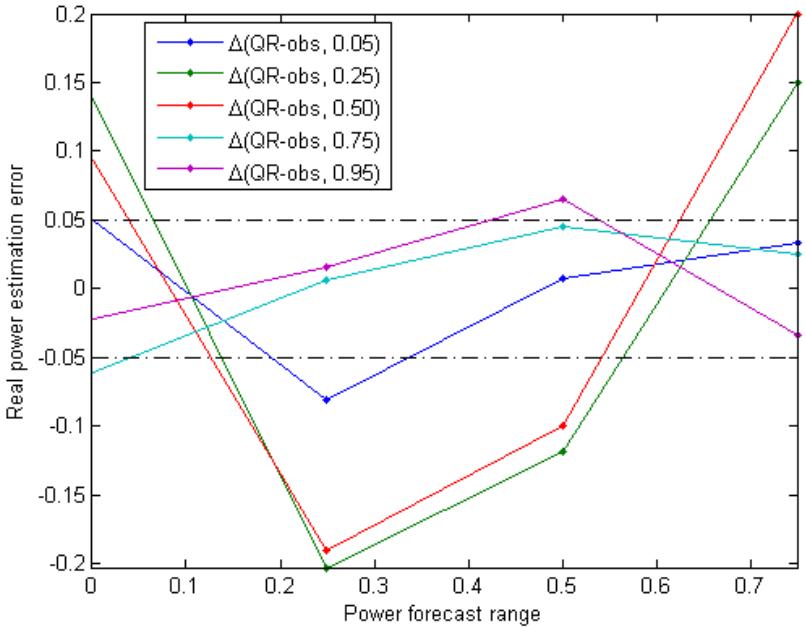

Figure 45: Difference of QR-based and observed production values when conditioned on predicted production

Table 10: Observed probabilities when conditioned on predicted production

$\hat{P}$ range	0.05	0.25	0.5	0.75	0.95
0.0 – 0.25	0	0.1111	0.4056	0.8111	0.9722
0.25 – 0.50	0.1310	0.4524	0.6905	0.7440	0.9345
0.50 – 0.75	0.0421	0.3684	0.6000	0.7053	0.8842
0.75 – 1.00	0.0167	0.1000	0.3000	0.7250	0.9833

8 Conclusions

This final Section of the thesis presents the summary of the work, accentuating on the most important issues and findings as well as the relevance of the obtained results. Some guidelines for future research and development are also discussed.

8.1 Summary

This thesis is written to serve as an introduction to wind power forecasting mainly for wind power companies operating in electricity wholesale markets. Given a set of existing wind power resources, wind power operators want to use them as efficiently and profitably as possible. In this pursuit, predicting the future output is one of their key interests. On the other hand, it is also a major challenge as wind prediction in general is one of the most arduous forecasting problems there is. This thesis has tried to evaluate and provide answers to the issues related to planning wind power production for the Nordic electricity markets.

We began by presenting some general characteristics of wind power such as its history, current use, and future perspectives, and then revealed some of the challenges related to its use. Such knowledge is essential before delving further into details. Furthermore, the intermittency of wind speed has been discussed, as well as its purpose in challenging the production planning.

All wind power is normally sold in electricity markets, whose internal structure poses limits to its production planning. Currently the markets expect the producer to communicate its next day's production the day before, and furthermore to engage in that amount of production. The producers also commit themselves to suffer from possible deviations. For the case of wind power, smaller or larger deviations are a rule rather than an exception, so that a producer must just accept a certain level of uncertainty due to omnipresent deviations. The level can be significantly lowered however by two methods. First of all, a wind power producer should use an advanced (and usually rather mathematically sophisticated) model as its forecasting tool. Secondly, the electricity markets provide the producers with a possibility to correct their initial commitments in intraday markets. By running very short-term predictions up to a few hours ahead, the final deviations can be dampened and the economical risk of profit loss minimized. Furthermore, we have elaborated on the possibilities of forming strategic bids of future wind power by taking into account the forecasting uncertainty.

In the last years, there has been extensive research in wind power forecasting methods. In this thesis, we have performed a rather thorough look into the scientific work in the field, and presented some of the most interesting and performant existing forecasting methods. One of the main goals has been to provide the reader an understanding on the applicability and limits of such methods, as well as trying to cover the whole large spectrum of different modeling approaches. It seems that although there are performant models that are in operational use, no model

has proved to be flawless in its prediction capabilities. By integrating different approaches and taking the best features from the individual modeling methods, wind power producers can nevertheless use their resources with a large degree of reliability and as operators in electricity markets hedge against consequential financial losses.

In order to assess the performance of the current forecasting procedure, several tests are performed. We have seen that although the forecasting model based on a turbine power curve performs reasonably well, there still remains at times large variations between the predicted and measured amount of production. Using a static power curve might not be optimal in the sense that it neglects systematic NWP (phase) biases and wind direction influence; in addition, it also does not offer the possibility to harness the strong autocorrelation of wind speed in short time horizons.

In order to better tackle the forecasting problems, increase the level of prediction accuracy and perform tests on some of the literature models, we have implemented several wind power forecasting approaches. Those models are then detailed and their implementation is discussed. From the results, the reader may note that using an advanced forecasting procedure generally pays off: the prediction errors diminish and the possibilities of estimating the prediction uncertainty are created.

Nevertheless, the nature of wind demands that our results must be understood applying only for the specific data used in evaluating different models. Although AI-based models theoretically can learn any types of relationships given any type of data, the structure of models may very well vary depending on the conditions of the wind farm site for instance. We thus restrict ourselves from giving any general recommendations of model types which would be universally optimal, as such models most likely do not globally exist. We insist however on the fact that based on scientific studies of ours and others, there are strong incentives for designing, using and maintaining a forecasting model that takes the best use of present-day state-of-the-art modeling techniques. In conclusion, it must be kept in mind that the models shall be continuously updated in order to account for seasonal variations and changes in the wind farm conditions.

8.2 Guidelines for future

The research concerning efficient, robust and operationally viable models in wind power forecasting and trading has provided the wind power companies with solutions aiding in the management of their wind resources. With these state-of-the-art models, a wind power company can significantly reduce the inherent uncertainty of its production level and thus improve its financial results. This has been shown to be true in this thesis as well; however, the research and development is far from complete. Considering the number of wind turbines shall increase significantly in Finland in the future, in turn there will be a need for even more efficient power forecasting. The same need is also shared by the transmission system operators, since a large amount of wind turbines feeding the electricity grids with a fluctuating supply of electricity can turn out to be a laborious problem with regard to managing

the grid balance.

Regarding the future needs of developing the forecasting model both theoretically and operationally, some guidelines warrant mentioning. Respecting these general issues one can design and develop wind power forecasting models to be even more accurate and informative.

Notably, the numerical weather prediction used in wind power forecasting is to some extent always biased and erronenous. The level of prediction errors depends strongly on the geographical conditions of the wind farm area; however, no single meteorological prediction has proven to be perfect. In order to quantify, understand and reduce this uncertainty, a wind power producer is strongly advised to test, and eventually procure multiple weather predictions. In this way a facile solution is obtained making standard and difficult forecasting conditions more easily identified, whereby producing an overall error reduction rate. Furthermore a closer collaboration between the operators of wind farms and the meteorologists is recommended, as the demands of numerical weather prediction tailored to the needs of wind power are very specific.

The wind speed prediction bias may also be accounted for by measuring wind speed from multiple locations near the wind farm. Using such sensor fusion techniques one can e.g. identify a wind burst coming from a certain direction and prepare the corrective actions concurrently. Particularly in the case of phase shift corrections, which are difficult to compute using only one wind speed measurement, this method would probably reduce forecasting errors significantly.

Secondly, the data collection systems in our case study wind farm were not of an acceptable level. There are no measurement masts for measuring the wind speed and direction at the case study location. It was therefore required to collect the data from individual turbines whose measurements are automatically biased, as the measurement is done after the turbine rotor plane. The same fact applies to measured production in addition to the index showing whether a turbine has been operative or in overhaul at a given moment. For model developers and in a later phase for operators as well, such information should be easily available, as the quality and applicability of such data is at the very heart of designing a performant forecasting system. All this leads to efforts in building and maintaining a control room in which all the necessary data concerning wind characteristics, turbine availability and future maintenance works, as well as electricity market conditions are easily available and ready to be employed in making efficient and reliable decisions on future production.

Thirdly, there naturally will be development in the power forecasting models themselves as well. One can see that the newest state-of-the-art models (e.g. support vector machines) perform better than the original classics (feed-forward neural networks) in the field of wind power forecasting. Such an improvement will hardly stop here. Nevertheless, looking at a simple illustration of measured wind and production plotted against each other, one can easily see that there is a rather large level of variation left unexplained by using wind speed as an explaining variable exclusively. Although it may be theoretically contradictory, employment of such variables as

wind direction, humidity, temperature etc. might indeed improve the forecasting results – provided that those individual forecasts are accurate. Certainly a single model is not enough to capture all the characteristics of wind speed / power fluctuations, so e.g. state-transitional and *Bayesian inference* models [71, 55] could be used in distinguishing different forecasting conditions. Also considering GARCH models (and its variants) in lieu of AR models in very short-term wind speed forecasting could be attempted.

In order to reduce the forecasting errors, one could also investigate the possibility of using e.g. 10 minute averages of wind speed and production time series. Using the data sampled at a higher frequency one could possibly employ information that is lost when formed into hourly averages. On the other hand, the averaging process itself also destroys some noise, so that care must be taken in filtering the high-frequency data appropriately. A good way of revealing characteristics at different frequency scales is to use wavelet analysis. Another maybe less known, albeit in a way more generic approach of decomposing signal into different levels of approximation, is a technique called *singular spectrum analysis* [3]. To the best of our knowledge based on published articles, it has not been used for wind power prediction purposes.

Participating in multiple electricity markets with large amounts of wind power capacity undoubtedly raises the interest in evaluating bidding opportunities which are optimized for hedging against the (expected) risk in the form of imbalance penalties. The research has already effectively provided theoretical bidding frameworks, yet there still is a lot to be considered as the characteristics of market imbalance prices for instance are rather complicated.

As theoretically shown in [32], which is also consistent in our findings concerning error growth with the forecasting look-ahead time, the nature of electricity day-ahead markets poses significant challenges for wind power. It is clear, that when having to predict 13-36 hours ahead for the next day's production, the producer is subject to significant variations and thus economical losses. Were the market foundations different in such a way that wind power producers would be allowed to predict their production later with a greater probability of actually producing the promised amount, the grid fluctuations would therefore diminish and the profits of wind power producers would rise. This would also have a positive effect on the security of electricity supply. Naturally there would also have to be enough electricity buyers willing to wait longer for wind power sellers to come to markets with their production, which in turn would increase the operating risk of buyers.

In conclusion, in order to achieve better forecasting results there are some essential points to be noted. The procurement of numerical weather predictions must be seen as one of the most fundamental factors affecting the forecasting results, whereupon wind power producers must carefully evaluate and purchase reliable meteorological data from several service providers. Data integrity and quality must be taken seriously as well, so that there are no outliers and all the measurements can be trusted and compared equally. Advanced forecasting models produce better results than merely using a simple turbine power curve; therefore time, effort and money must

be allocated for designing a performant forecasting system, as well as educating its operators. With such a system providing informative, accurate and reliable predictions, operators can take effective actions in order to utilize wind resources with maximum effectiveness.

References

[1] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.

[2] S. A. Akdag and A. Dinler. A new method to estimate Weibull parameters for wind energy applications. *Energy Conversion and Management*, 50(7):1761–1766, 7/ 2009. doi: 10.1016/j.enconman.2009.03.020.

[3] C. Aldrich and M. Barkhuizen. Process system identification strategies based on the use of singular spectrum analysis. *Minerals Engineering*, 16(9):815–826, 9/ 2003. doi: 10.1016/S0892-6875(03)00203-6.

[4] N. Amjady and F. Keynia. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy*, 34(1):46–57, 1/ 2009. doi: 10.1016/j.energy.2008.09.020.

[5] J. L. Angarita, J. Usaola, and J. Martínez-Crespo. Combined hydro-wind generation bids in a pool-based electricity market. *Electric Power Systems Research*, 79(7):1038–1046, 7/ 2009. doi: 10.1016/j.epsr.2009.01.002.

[6] European Wind Energy Association. *Wind Energy - The Facts*. Earthscan Ltd, 2009.

[7] R. J. Barthelmie, F. Murray, and S. C. Pryor. The economic benefit of short-term forecasting for wind energy in the UK electricity market. *Energy Policy*, 36(5):1687–1696, 5/ 2008. doi: 10.1016/j.enpol.2008.01.027.

[8] J. P. Barton and D. G. Infield. Energy storage and its use with intermittent renewable energy. *IEEE Transactions on Energy Conversion*, 19(2):441–448, 2004.

[9] J. M. Bates and C. W. J. Granger. The combination of forecasts. *OR*, 20(4):451–468, 1969.

[10] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 2000.

[11] C. De Boor. *A Practical Guide to Splines*. Springer, Berlin, 1978.

[12] G. E. P. Box, G. C. Reinsel, and G. M. Jenkins. *Time series analysis: forecasting and control*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[14] E. Cadenas and W. Rivera. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. *Renewable Energy*, 34(1):274–278, 1/ 2009. doi: 10.1016/j.renene.2008.03.014.

[15] J. A. Carta and D. Mentado. A continuous bivariate model for wind power density and wind turbine energy output estimations. *Energy Conversion and Management*, 48(2):420–432, 2/ 2007. doi: 10.1016/j.enconman.2006.06.019.

[16] M. Clements and D. Hendry. *A Companion to Economic Forecasting*. Wiley-Blackwell, 2002.

[17] European Commission. The EU climate and energy package, 2007. Accessed April 6, 2010. Available online: http://ec.europa.eu/environment/climat/climate_action.htm.

[18] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725–1744, 8/ 2008. doi: 10.1016/j.rser.2007.01.015.

[19] A. Crespo, J. Hernández, and S. Frandsen. Survey of modelling methods for wind turbine wakes and wind farms. *Wind Energy*, 2(1):1–24, 1999.

[20] R. Durrett. *Essentials of Stochastic Processes*. Springer, 2001.

[21] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[22] M. Esponiza, T. Falck, J. Suykens, and B. De Moor. Time series prediction using LS-SVMs. Technical report, 2008. Accessed February 3, 2010. Available online: http://www.mafy.lut.fi/timeseries/ESTSP/PDF/16.pdf.

[23] J.-C. Bouette et al. Wind in Ireland: long memory or seasonal effect? *Stochastic Environmental Research and Risk Assessment*, 20(3):141–151, 4/ 2006. doi: 10.1007/s00477-005-0029-y.

[24] D. L. Faria, R. Castro, C. Philippart, and A. Gusmao. Wavelets pre-filtering in wind speed prediction. In *International Conference on Power Engineering, Energy and Electrical Drives*, pages 168–173, 2009.

[25] Fingrid. Press releases: Nordic electricity market expanding to the Baltic countries - will price spikes remain in history?, 3/ 2010. Accessed March 24, 2010. Available online: http://www.fingrid.fi/portal/in_english/news_and_releases/press_releases.

[26] J. Flores, R. Loaeza, H. Rodriguez, and E. Cadenas. Wind speed forecasting using a hybrid neural-evolutive approach, 2009. doi: 10.1007/978-3-642-05258-3_53.

[27] U. Focken and M. Lange. Previento - regional wind power prediction with risk control. Technical report, 4/ 2008. Accessed January 28, 2010. Available online: http://ehf.uni-oldenburg.de/wind/download/paris2002/Previento-Risk-Control.pdf.

[28] Riso National Laboratory for Sustainable Energy. Wind atlases of the world. Accessed February 8, 2010. Available online: http://www.windatlas.dk.

[29] G. Giebel, R. Brownsword, and G. Kariniotakis. The state-of-the-art in short-term prediction of wind power. Technical report, 2003. Accessed February 24, 2010. Available online: http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf.

[30] I. Goethals, J. Suykens, K. Pelckmans, and B. De Moor. NARX identification of Hammerstein models using least squares support vector machines. Technical report, 2004. Accessed March 2, 2010. Available online: ftp://ftp.esat.kuleuven.ac.be/pub/SISTA//kpelckma/nolcos2004_ig.pdf.

[31] K. W. Hedman and G. B. Sheblé. Comparing hedging methods for wind power: Using pumped storage hydro units vs. options purchasing. In *International Conference on Probabilistic Methods Applied to Power Systems*, pages 1–6, 2006.

[32] H. Holttinen. Optimal electricity market for wind power. *Energy Policy*, 33(16):2052–2063, 11/ 2005. doi: 10.1016/j.enpol.2004.04.001.

[33] Y. Y. Hong and C. P. Wu. Short-term wind power and speed forecasting using association rules and adaptive network-based fuzzy inference system. *International Journal of Electrical Engineering*, 16(2):151–158, 2009.

[34] D. Hou, E. Kalnay, and K. K. Droegemeier. Objective verification of the SAMEX '98 ensemble forecasts. *Monthly Weather Review*, pages 73–91, 2001.

[35] D. Hunter. Matlab code for quantile regression. Accessed February 20, 2010. Available online: http://www.stat.psu.edu/%7edhunter/code/qrmatlab/.

[36] Finnish Meteorological Institute. Finnish wind atlas, 2010. Accessed January 15, 2010. Available online: http://www.tuuliatlas.fi/modelling/index.html.

[37] J. S. R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3):665–685, 1993.

[38] L. Järvenpää. Sähkömarkkinoiden toiminta ja hinnanmuodostus: Pohjoismaisten Nordpool-markkinoiden tarkastelua, 2010. Bachelor's Thesis. University of Helsinki, Faculty of Social Sciences. In Finnish.

[39] J. Juban, N. Siebert, and G. N. Kariniotakis. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In *IEEE Lausanne Power Tech*, pages 683–688, 2007.

[40] O. Kauppi. A model of imperfect dynamic competition, 2009. Doctoral Dissertation. Helsinki School of Economics.

[41] R. G. Kavasseri and K. Seetharaman. Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy*, 34(5):1388–1393, 5/ 2009. doi: 10.1016/j.renene.2008.09.006.

[42] A. A. Khan and M. Shahidehpour. One day ahead wind speed forecasting using wavelets. In *IEEE Power Systems Conference and Exposition*, pages 1–5, 2009.

[43] M. Lange. On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors. *Journal of Solar Energy Engineering*, 127(2):177–184, 5/ 2005.

[44] M. Lange and U. Focken. New developments in wind energy forecasting. In *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–8, 2008.

[45] C. Lei and L. Ran. Short-term wind speed forecasting model for wind farm based on wavelet decomposition. In *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 2525–2529, 2008.

[46] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915–920, 5/ 2009. doi: 10.1016/j.rser.2008.02.002.

[47] G. Li and J. Shi. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy*, In Press, Corrected Proof.

[48] U. Linnet. Tools supporting wind energy trade in deregulated markets, 2005. Master's thesis. Technical University of Denmark.

[49] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(12):2348–2362, 12/ 2008. doi: 10.1016/j.jweia.2008.03.013.

[50] Z. Lu. Analyse des processus longue mémoire stationnaires et non-stationnaires: Estimations, applications et prévisions, 2009. Doctoral Dissertation. Ecole Normale Supérieure de Cachan.

[51] J. J. Lucia and E. S. Schwartz. Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Review of Derivatives Research*, 5(1):5–50, 1/ 2002. doi: 10.1023/A:1013846631785.

[52] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[53] J. Matevosyan, M. Olsson, and L. Söder. Hydropower planning coordinated with wind power in areas with congestion problems for trading on the spot and the regulating market. *Electric Power Systems Research*, 79(1):39–48, 1/ 2009. doi: 10.1016/j.epsr.2008.05.019.

[54] J. Matevosyan and L. Söder. Minimization of imbalance cost trading wind power on the short-term power market. *IEEE Transactions on Power Systems*, 21(3):1396–1404, 2006.

[55] M. S. Miranda and R. W. Dunn. One-hour-ahead wind speed prediction using a Bayesian methodology. In *IEEE Power Engineering Society General Meeting*, 2006.

[56] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz. Short-term trading for a wind power producer. *IEEE Transactions on Power Systems*, In Press, 2010.

[57] H. A. Nielsen, H. Madsen, and T. S. Nielsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, 9(1-2):95–108, 2006.

[58] H. A. Nielsen, T. S. Nielsen, H. Madsen, M. J. San Isidro Pindado, and I. Marti. Optimal combination of wind power forecasts. *Wind Energy*, 10(5):471–482, 2007.

[59] NordPool, Price Calculation Principles. Accessed February 2, 2010.
Available online: http://www.nordpoolspot.com/trading/The_Elspot_market.

[60] Finnish Ministry of Employment and the Economy. Hallituksen esitys eduskunnalle laiksi uusiutuvilla energialähteillä tuotetun sähkön tuotantotuesta, 2010. Accessed April 7, 2010.
Available online: http://www.tem.fi/files/26294/HEluonnos_110310_.pdf. In Finnish.

[61] M. Olsson and L. Söder. Generation of regulating power price scenarios. In *International Conference on Probabilistic Methods Applied to Power Systems*, pages 26–31, 2004.

[62] European Transmission System Operators. Report on balance management harmonisation and integration, 2007. Accessed February 18, 2010.
Available online: http://www.marketcoupling.com/downloads/downloads.

[63] K. Pelckmans and J. Suykens. LS-SVMlab: a MATLAB/C toolbox for Least Squares Support Vector Machines. Technical report, 2002. Accessed February 26, 2010.
Available online: http://www.esat.kuleuven.be/sista/lssvmlab/.

[64] P. Pinson. Estimation of the uncertainty in wind power forecasting, 2006. Doctoral Dissertation. Ecole des Mines de Paris.

[65] P. Pinson, C. Chevallier, and G. N. Kariniotakis. Trading wind generation from short-term probabilistic forecasts of wind power. *IEEE Transactions on Power Systems*, 22(3):1148–1156, 2007.

[66] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12(1):51–62, 2009.

[67] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. *Wind Energy*, 10(6):497–516, 2007.

[68] P. Pinson, H. A. Nielsen, H. Madsen, and G. Kariniotakis. Skill forecasting from ensemble predictions of wind power. *Applied Energy*, 86(7-8):1326–1334, 2009. doi: 10.1016/j.apenergy.2008.10.009.

[69] S. Pöyhönen. Support vector machine based classification in condition monitoring of induction motors, 2004. Doctoral Dissertation. Helsinki University of Technology.

[70] M. Rajala. Integroituvat Itämeren sähkömarkkinat, 2008. Master's thesis. Lappeenranta University of Technology. In Finnish.

[71] G. Reikard. Using temperature and state transitions to forecast wind speed. *Wind Energy*, 11(5):431–443, 2008.

[72] I. Sanchez. Adaptive combination of forecasts with application to wind energy. *International Journal of Forecasting*, 24(4):679–693, 12/ 2008. doi: 10.1016/j.ijforecast.2008.08.008.

[73] H. Savijärvi. Numeerinen meteorologia I. Technical report, University of Helsinki, 2007. In Finnish.

[74] A. F. Sheta and K. De Jong. Time-series forecasting using GA-tuned radial basis functions. *Information Sciences*, 133(3-4):221–228, 4/ 2001.

[75] G. Sideratos and N. D. Hatziargyriou. An advanced statistical method for wind power forecasting. *IEEE Transactions on Power Systems*, 22(1):258–265, 2007.

[76] S. N. Singh and I. Erlich. Strategies for wind power trading in competitive electricity markets. *IEEE Transactions on Energy Conversion*, 23(1):249–256, 2008.

[77] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 8/ 2004. doi: 10.1023/B:STCO.0000035301.49549.88.

[78] D. Svozil, V. Kvasnicka, and J. Pospichal. Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1):43–62, 11/ 1997.

[79] J. L. Torres, A. Garcia, M. De Blas, and A. De Francisco. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, 79(1):65–77, 7/ 2005. doi: 10.1016/j.solener.2004.09.013.

[80] J. Usaola and M. A. Moreno. Optimal bidding of wind energy in intraday markets. In *6th International Conference on the European Energy Market*, pages 1–7, 2009.

[81] L. Vandezande, L. Meeus, R. Belmans, M. Saguan, and J.-M. Glachant. Well-functioning balancing markets: A prerequisite for wind power integration. *Energy Policy*, In Press, Corrected Proof. doi: 10.1016/j.enpol.2009.07.034.

[82] V. N. Vapnik. *The Nature of Statistical Learning*. Springer, 2000.

[83] H. R. Varian. *Microeconomic Analysis*. W.W. Norton & Co., 1992.

[84] I. Vehviläinen. Applying mathematical finance tools to the competitive Nordic electricity market, 2004. Doctoral Dissertation. Helsinki University of Technology.

[85] M. Ventosa, A. Baillo, A. Ramos, and M. Rivier. Electricity market modeling trends. *Energy Policy*, 33(7):897–913, 5/ 2005. doi: 10.1016/j.enpol.2003.10.013.

[86] Y.K. Wu and J.S. Hong. A literature review of wind forecasting technology in the world. In *IEEE Lausanne Power Tech*, pages 504–509, 2007.

[87] A. Yamaguchi, T. Ishihara, and Y. Fujino. Experimental study of the wind flow in a coastal region of Japan. *Journal of Wind Engineering and Industrial Aerodynamics*, 91(1-2):247–264, 1/ 2003. doi: 10.1016/S0167-6105(02)00349-5.